

Universidad de Alcalá

Escuela Politécnica Superior

Grado en Ingeniería en Electrónica y Automática Industrial



Trabajo Fin de Grado

SIMULACIÓN DE SEÑALES EN PLASMAS DE FUSIÓN NUCLEAR
MEDIANTE TÉCNICAS DE REGRESIÓN PARAMÉTRICA

Autor: Pablo Gilaberte López

Tutor: Augusto Pereira González

2021

UNIVERSIDAD DE ALCALÁ

Escuela Politécnica Superior

Grado en Ingeniería en Electrónica y Automática Industrial

Trabajo Fin de Grado

SIMULACIÓN DE SEÑALES EN PLASMAS DE FUSIÓN NUCLEAR
MEDIANTE TÉCNICAS DE REGRESIÓN PARAMÉTRICA

Autor: Pablo Gilaberte López

Tutor: Augusto Pereira González

TRIBUNAL:

Presidente: Cristina Alén Cordero

Vocal 1º: Juan Antonio Martínez Rojas

Vocal 2º: Augusto Pereira González

FECHA: 25 de Junio de 2021

Simulación de señales en plasmas de fusión nuclear mediante técnicas de regresión paramétrica

Autor: Pablo Gilaberte López

Tutor: Augusto Pereira González

Departamento: Teoría de la Señal y Comunicaciones

Titulación: Grado en Ingeniería Electrónica y Automática Industrial

Palabras clave: Regresión simbólica, ley de potencia, ley de escala, modelo de regresión.

Resumen

Aplicación y comparación de modelos de regresión predictiva utilizando cinco señales de evolución temporal que recogen distintos parámetros físicos del dispositivo experimental de fusión nuclear TJ-II del CIEMAT. Se han comparado dos técnicas de regresión paramétrica diferentes. Por un lado, se ha utilizado la regresión simbólica en combinación con los algoritmos genéticos y el segundo de los métodos ha consistido en el uso de la regresión estadística en forma de leyes de escala o también conocidas como leyes de potencia. Finalmente, para seleccionar el modelo predictivo más fidedigno se ha utilizado el mínimo error en la comparación entre la curva real original de la señal y la simulada sintéticamente mediante los dos modelos anteriormente descritos.

Simulation of signals in nuclear fusion plasmas using parametric regression techniques

Author: Pablo Gilaberte López

Tutor: Augusto Pereira González

Department: Signal Theory and Communications

Degree: Degree in Industrial Electronics and Automation Engineering

Key words: Symbolic regression, power law, scaling law, regression model.

Abstract

Application and comparison of predictive regression models using five-time evolution signals that collect different physical parameters of the CIEMAT TJ-II nuclear fusion experimental device. Two different parametric regression techniques have been compared. On the one hand, symbolic regression has been used in combination with genetic algorithms and the second method consisted in the use of statistical regression in the form of scaling laws or also known as power laws. Finally, to select the most reliable predictive model, the minimum error in the comparison between the original real curve of the signal and the one simulated synthetically by means of the two models described above was used.

INDICE

Estructura de la memoria.....	9
INTRODUCCIÓN	10
Objetivos	12
Regresión Paramétrica	13
Regresión Simbólica	13
Leyes de Potencia.....	15
Leyes semejantes	16
Métodos analíticos	16
Metodología de Procesos.....	27
Desarrollo Regresión Simbólica.....	29
Desarrollo Leyes de Potencia	35
Resultados prácticos	45
Calentamiento tipo 1.....	45
Calentamiento tipo 2.....	50
Calentamiento tipo 3.....	55
Calentamiento tipo 4.....	60
Calentamiento tipo 5.....	65
Calentamiento tipo 6.....	70
Calentamiento tipo 7.....	75
Conclusiones	81
Bibliografía	87

TABLA DE ILUSTRACIONES

ILUSTRACIÓN 1.DISTRIBUCIÓN EN LEYES DE POTENCIA.....	15
ILUSTRACIÓN 2.GRÁFICA ORDINARY LEAST SQUARE	16
ILUSTRACIÓN 3.GRAFICA OLS.....	17
ILUSTRACIÓN 4.GRÁFICA RIDGE REGRESSION	18
ILUSTRACIÓN 5.FACTOR K DE RIDGE REGRESSION.	19
ILUSTRACIÓN 6.GRAFICA DE TLS.	20
ILUSTRACIÓN 7.GRÁFICA DE ROBUST REGRESSION	21
ILUSTRACIÓN 8.RECTAS DE DISTINTOS TIPOS DE ROBUST REGRESSION	22
ILUSTRACIÓN 9.MEMORIA EN USO POR BASE DE DATOS.....	23
ILUSTRACIÓN 10.INICIO DE BASE DE DATOS EN MATLAB.....	23
ILUSTRACIÓN 11.FIN DE BASE DE DATOS EN MATLAB.....	24
ILUSTRACIÓN 12.COMPARATIVA TRANSFORMADA DE HAAR	25
ILUSTRACIÓN 13.DESCARGAS POR TIPO DE CALENTAMIENTO	26
ILUSTRACIÓN 14.LISTADO DE DESCARGAS EN BASE DE DATOS NORMALIZADA.....	27
ILUSTRACIÓN 15.FUNCION FIND PREVIOUS DISCHARGES.M	28
ILUSTRACIÓN 16.RESULTADO DE DESCARGAS SIMILARES.....	28
ILUSTRACIÓN 17.VALORES DE CADA VARIABLE CON SU RESPECTIVA DESCARGA.....	29
ILUSTRACIÓN 18.BASE DE DATOS EN EUREQA	30
ILUSTRACIÓN 19. DEFINICIÓN DE VARIABLES DEPENDIENTE E INDEPENDIENTES AL DEFINIR EL MODELO.....	31
ILUSTRACIÓN 20.TIEMPO DE DURACIÓN EN OBTENER EL MODELO	31
ILUSTRACIÓN 21.CURVA ERROR VS TIEMPO	32
ILUSTRACIÓN 22.ECUACIONES DEL MODELO	32
ILUSTRACIÓN 23.CURVA DEL MODELO VS CURVA ESTIMADA	33
ILUSTRACIÓN 24.CURVA ERROR VS COMPLEJIDAD	33
ILUSTRACIÓN 25.ECUACIÓN DEL MODELO FINAL	33
ILUSTRACIÓN 26. FUNCIÓN RUNEQUATION.M	34
ILUSTRACIÓN 27.FUNCIÓN OLS	36
ILUSTRACIÓN 28.FUNCIÓN RIDGE REGRESSION	38
ILUSTRACIÓN 29.FUNCION TLS.....	39
ILUSTRACIÓN 30.FUNCIÓN RTLS.....	41
ILUSTRACIÓN 31.FUNCIÓN ROBUST REGRESSION.....	44
ILUSTRACIÓN 32.SCRIPT FINAL	44
ILUSTRACIÓN 33. TE REGRESIÓN SIMBÓLICA CALENTAMIENTO 1	45
ILUSTRACIÓN 34. TE LEYES DE POTENCIA CALENTAMIENTO 1	46
ILUSTRACIÓN 35. N REGRESIÓN SIMBÓLICA CALENTAMIENTO 1	46
ILUSTRACIÓN 36. N LEYES DE POTENCIA CALENTAMIENTO 1	47
ILUSTRACIÓN 37. WP REGRESIÓN SIMBÓLICA CALENTAMIENTO 1.....	47
ILUSTRACIÓN 38. WP LEYES DE POTENCIA CALENTAMIENTO 1	48
ILUSTRACIÓN 39. IP REGRESIÓN SIMBÓLICA CALENTAMIENTO 1	48
ILUSTRACIÓN 40. IP LEYES DE POTENCIA CALENTAMIENTO 1	49
ILUSTRACIÓN 41. HA REGRESIÓN SIMBÓLICA CALENTAMIENTO 1.....	49
ILUSTRACIÓN 42. HA LEYES DE POTENCIA CALENTAMIENTO 1	50
ILUSTRACIÓN 43. TE REGRESIÓN SIMBÓLICA CALENTAMIENTO 2	50
ILUSTRACIÓN 44. TE LEYES DE POTENCIA CALENTAMIENTO 2	51
ILUSTRACIÓN 45. N REGRESIÓN SIMBÓLICA CALENTAMIENTO 2	51
ILUSTRACIÓN 46. N LEYES DE POTENCIA CALENTAMIENTO 2	52
ILUSTRACIÓN 47. WP REGRESIÓN SIMBÓLICA CALENTAMIENTO 2.....	52
ILUSTRACIÓN 48. WP LEYES DE POTENCIA CALENTAMIENTO 2	53
ILUSTRACIÓN 49. IP REGRESIÓN SIMBÓLICA CALENTAMIENTO 2	53

ILUSTRACIÓN 50. IP LEYES DE POTENCIA CALENTAMIENTO 2	54
ILUSTRACIÓN 51. HA REGRESIÓN SIMBÓLICA CALENTAMIENTO 2	54
ILUSTRACIÓN 52. HA LEYES DE POTENCIA CALENTAMIENTO 2	55
ILUSTRACIÓN 53. TE REGRESIÓN SIMBÓLICA CALENTAMIENTO 3	55
ILUSTRACIÓN 54. TE LEYES DE POTENCIA CALENTAMIENTO 3	56
ILUSTRACIÓN 55. N REGRESIÓN SIMBÓLICA CALENTAMIENTO 3	56
ILUSTRACIÓN 56. N LEYES DE POTENCIA CALENTAMIENTO 3	57
ILUSTRACIÓN 57. WP REGRESIÓN SIMBÓLICA CALENTAMIENTO 3	57
ILUSTRACIÓN 58. WP LEYES DE POTENCIA CALENTAMIENTO 3	58
ILUSTRACIÓN 59. IP REGRESIÓN SIMBÓLICA CALENTAMIENTO 3	58
ILUSTRACIÓN 60. IP LEYES DE POTENCIA CALENTAMIENTO 3	59
ILUSTRACIÓN 61. HA REGRESIÓN SIMBÓLICA CALENTAMIENTO 3	59
ILUSTRACIÓN 62. HA LEYES DE POTENCIA CALENTAMIENTO 3	60
ILUSTRACIÓN 63. TE REGRESIÓN SIMBÓLICA CALENTAMIENTO 4	60
ILUSTRACIÓN 64. TE LEYES DE POTENCIA CALENTAMIENTO 4	61
ILUSTRACIÓN 65. N REGRESIÓN SIMBÓLICA CALENTAMIENTO 4	61
ILUSTRACIÓN 66. N LEYES DE POTENCIA CALENTAMIENTO 4	62
ILUSTRACIÓN 67. WP REGRESIÓN SIMBÓLICA CALENTAMIENTO 4	62
ILUSTRACIÓN 68. WP LEYES DE POTENCIA CALENTAMIENTO 4	63
ILUSTRACIÓN 69. IP REGRESIÓN SIMBÓLICA CALENTAMIENTO 4	63
ILUSTRACIÓN 70. IP LEYES DE POTENCIA CALENTAMIENTO 4	64
ILUSTRACIÓN 71. HA REGRESIÓN SIMBÓLICA CALENTAMIENTO 4	64
ILUSTRACIÓN 72. HA LEYES DE POTENCIA CALENTAMIENTO 4	65
ILUSTRACIÓN 73. TE REGRESIÓN SIMBÓLICA CALENTAMIENTO 5	65
ILUSTRACIÓN 74. N LEYES DE POTENCIA CALENTAMIENTO 5	66
ILUSTRACIÓN 75. N REGRESIÓN SIMBÓLICA CALENTAMIENTO 5	66
ILUSTRACIÓN 76. TE LEYES DE POTENCIA CALENTAMIENTO 5	67
ILUSTRACIÓN 77. WP REGRESIÓN SIMBÓLICA CALENTAMIENTO 5	67
ILUSTRACIÓN 78. WP LEYES DE POTENCIA CALENTAMIENTO 5	68
ILUSTRACIÓN 79. IP REGRESIÓN SIMBÓLICA CALENTAMIENTO 5	68
ILUSTRACIÓN 80. IP LEYES DE POTENCIA CALENTAMIENTO 5	69
ILUSTRACIÓN 81. HA REGRESIÓN SIMBÓLICA CALENTAMIENTO 5	69
ILUSTRACIÓN 82. HA LEYES DE POTENCIA CALENTAMIENTO 5	70
ILUSTRACIÓN 83. TE REGRESIÓN SIMBÓLICA CALENTAMIENTO 6	70
ILUSTRACIÓN 84. TE LEYES DE POTENCIA CALENTAMIENTO 6	71
ILUSTRACIÓN 85. N REGRESIÓN SIMBÓLICA CALENTAMIENTO 6	71
ILUSTRACIÓN 86. N LEYES DE POTENCIA CALENTAMIENTO 6	72
ILUSTRACIÓN 87. WP REGRESIÓN SIMBÓLICA CALENTAMIENTO 6	72
ILUSTRACIÓN 88. WP LEYES DE POTENCIA CALENTAMIENTO 6	73
ILUSTRACIÓN 89. IP REGRESIÓN SIMBÓLICA CALENTAMIENTO 6	73
ILUSTRACIÓN 90. TE LEYES DE POTENCIA CALENTAMIENTO 6	74
ILUSTRACIÓN 91. HA REGRESIÓN SIMBÓLICA CALENTAMIENTO 6	74
ILUSTRACIÓN 92. HA LEYES DE POTENCIA CALENTAMIENTO 6	75
ILUSTRACIÓN 93. TE REGRESIÓN SIMBÓLICA CALENTAMIENTO 7	75
ILUSTRACIÓN 94. TE LEYES DE POTENCIA CALENTAMIENTO 7	76
ILUSTRACIÓN 95. N REGRESIÓN SIMBÓLICA CALENTAMIENTO 7	76
ILUSTRACIÓN 96. N LEYES DE POTENCIA CALENTAMIENTO 7	77
ILUSTRACIÓN 97. WP REGRESIÓN SIMBÓLICA CALENTAMIENTO 7	77
ILUSTRACIÓN 98. WP LEYES DE POTENCIA CALENTAMIENTO 7	78
ILUSTRACIÓN 99. IP REGRESIÓN SIMBÓLICA CALENTAMIENTO 7	78
ILUSTRACIÓN 100. IP LEYES DE POTENCIA CALENTAMIENTO 7	79
ILUSTRACIÓN 101. HA REGRESIÓN SIMBÓLICA CALENTAMIENTO 7	79
ILUSTRACIÓN 102. HA LEYES DE POTENCIA CALENTAMIENTO 7	80

ILUSTRACIÓN 103.COMPARATIVA CALENTAMIENTO 1	81
ILUSTRACIÓN 104. COMPARATIVA CALENTAMIENTO 2.....	82
ILUSTRACIÓN 105. COMPARATIVA CALENTAMIENTO 3.....	82
ILUSTRACIÓN 106. COMPARATIVA CALENTAMIENTO 4.....	83
ILUSTRACIÓN 107. COMPARATIVA CALENTAMIENTO 5.....	83
ILUSTRACIÓN 108. COMPARATIVA CALENTAMIENTO 6.....	84
ILUSTRACIÓN 109. COMPARATIVA CALENTAMIENTO 7.....	84
ILUSTRACIÓN 110.DOMINACIÓN POR GRÁFICAS	85
ILUSTRACIÓN 111. ORDEN DE LOGROS DE LEYES DE POTENCIA	85
TABLA 1. RESULTADOS DEL ANÁLISIS COMPARATIVO.....	86

Estructura de la memoria

Dicho documento se ha organizado en once secciones.

En la primera sección, se realiza una breve introducción acerca de cuáles son las bases del proyecto, explicando sus fundamentos y orígenes iniciales a la hora de estudiar dichas señales. También se realiza una exposición acerca de la finalidad de dicho análisis, así como sus posibles implicaciones en futuras situaciones.

En la segunda sección se procede a la explicación del fundamento de las técnicas de regresión paramétrica, razonando el empleo de dicho método de análisis, en vez de usar otro procedimiento. Adicionalmente, se explican los dos métodos empleados basados en la regresión paramétrica, estos son la regresión simbólica y las leyes de potencia.

Posteriormente en la tercera sección, se exponen las características y como se encuentran organizadas las bases de datos, a partir de las cuales se toman todos los valores con los que posteriormente se realiza el análisis.

En cuarto lugar, se explica la preparación y preprocesamiento de datos, en el cual se muestra toda la serie de pasos llevados a cabo para poder tratar los valores de la base de datos de una forma correcta, ya que de forma original sería imposible operar con ellos.

En el quinto apartado, se muestra el método que se ha llevado a cabo y los distintos procesos, así como las funciones a nivel de programación que se han empleado.

En el sexto y séptimo capítulo, se muestra el análisis mediante regresión simbólica y leyes de potencia, respectivamente. En dichos puntos se explica que proceso se ha llevado a cabo hasta obtener el modelo matemático definitivo que más se ha aproximado a la realidad.

En la octava sección, se procede a mostrar el resultado que más se aproxima a la realidad procedente de cada uno de los dos métodos empleados.

En el noveno capítulo, se realiza una serie de conclusiones en las cuales se expone cual es el mejor método para obtener el modelo matemático más aproximado a la realidad en caso de ser necesario aplicarlo.

En el décimo capítulo se muestra toda la bibliografía empleada, que ha sido empleada tanto para el proceso de programación, como para el tratamiento de las bases de datos o la toma de conclusiones.

Por último, se muestran todas gráficas que se han aproximado a la realidad en menor grado, presentando errores por mínimos cuadrados superiores a los mostrados en las gráficas del apartado 6 y 7, así como todos los scripts programados para el desarrollo de todas las funciones empleadas.

INTRODUCCIÓN

Antes de iniciar la explicación acerca de los distintos puntos que se desarrollarán con gran detalle a lo largo de toda la memoria, se ha de realizar una contextualización de los hechos, con el fin de que todo lo descrito posteriormente no carezca de sentido.

En primer lugar, se debe de mencionar el dispositivo experimental de fusión termonuclear TJ-II. [Pereira,2010]. Dicho elemento tecnológico es la base del experimento, a través del cual se emplea gran información proveniente de él. Dichos datos son calificados como señales de evolución temporal, los cuales se han obtenido en pruebas realizadas a lo largo de mucho tiempo y almacenadas en dicho dispositivo. Del mismo modo, dichos datos son almacenados con el objetivo de que se perfeccionen los métodos analíticos mediante los cuales se obtenga una gran cantidad de información que permita descubrir patrones o combinaciones difíciles de ser encontradas. Por otra parte, todo el tratamiento de los datos procedentes de dicho componente da paso a que se pueda reproducir, del mismo modo de forma más rápida y eficaz en el proyecto ITER. Dicho programa está diseñado con la finalidad de demostrar a la comunidad científica y a la sociedad, que la fusión nuclear presenta una gran viabilidad como fuente de energía a gran escala. Como puntos importantes a destacar, se menciona la capacidad de producir una potencia de 500MW, también ser el primer dispositivo de fusión capaz de obtener una ganancia neta de energía, es decir, el sistema será capaz de producir una mayor cantidad de energía que la que absorbe para su funcionamiento. Por último, demostrar que dicho dispositivo presenta una gran seguridad sin causar daños a la naturaleza. [ITER, 2015]

Otro aspecto al que se debe de hacer mención en el contexto de dicho proyecto es procesamiento de datos masivos. Actualmente, las técnicas empleadas en el denominado big data [Powerdata,2020], son la base de cualquier estudio encargado de realizar un análisis en el cual se tratan cantidades desmesuradamente elevadas de datos. De este modo, lo que se consigue mediante estos métodos es que se logre relacionar ciertos datos que a simple vista son imperceptibles por el ser humano, pero que realmente permiten obtener patrones que desvelan informaciones muy importantes. De este modo, el análisis de los datos proporcionados por el dispositivo TJ-II presenta lo descrito previamente, ya que con su extensa base de datos, se deben de aplicar técnicas de preprocesamiento, así como algoritmos o transformadas matemáticas, con la finalidad de no alterar las características esenciales de dichos valores, pero reduciendo el tamaño total de la base de tal modo que se hayan normalizado los valores, y sea mucho más cómodo trabajar con ellos y que se puedan encontrar correlaciones internas sin demorarse demasiado en el tiempo.

En base a lo explicado, cabe destacar que cuando se pretende realizar una investigación con el fin de optimizar el control de la reacción de fusión, debe de presentarse una gran cantidad de información reflejada mediante gran diversidad de parámetros, con el fin de que el estudio sea lo más completo posible, y ayude a encontrar la máxima correlación posible entre los datos. Todos estos parámetros se consiguen mediante el uso de sensores, los cuales son los encargados de captar cualquier magnitud física y transformarla en una tensión eléctrica, donde a continuación se realiza una fase de adaptación, en la cual se realiza la conversión de dicha señal analógica en una señal digital.

También se debe de destacar el empleo de distintos entornos en lo referente a la programación, dado que el preprocesamiento de datos y los métodos llevados a cabo se realizan de forma

remota. De este modo, se muestra un aspecto muy visible de facilidad a la hora de realizar estudios referentes a este ámbito, y de que no exista una necesidad de disponer de una tecnología elevada para que se lleguen a obtener grandes conclusiones. Todo el análisis es posible realizarlo mediante diferentes programas informáticos, los cuales dan la posibilidad de visualizar mediante gráficas, obtener modelos matemáticos, e incluso emplear el resultado final de diferentes softwares en otro completamente distinto, complementándose y obteniendo conclusiones de gran interés.

Objetivos

El análisis de regresión estadística nos aporta información de cómo matemáticamente se comporta un sistema a partir de unos datos de entrada. Mediante la obtención de una ecuación matemática se pueden modelar los datos obtenidos por un sistema físico con el mínimo error posible. El mejor modelo matemático se ajusta al mejor modelo físico del sistema cuando en los datos de partida no existen errores en los mismos. En la práctica, en la mayoría de los sistemas físicos y de ingeniería esto no ocurre así. Las variables de medida suelen estar contaminadas por diferentes perturbaciones ambientales y electromagnéticas. Además, los componentes mecánicos de un sistema se van desajustando progresivamente, produciendo datos mal acondicionados. En los dispositivos experimentales de fusión nuclear sucede lo mismo. En estos dispositivos, se investiga el comportamiento del plasma, isótopos de hidrógeno ionizados a muy elevadas temperaturas, mediante su confinamiento con trampas magnéticas. Multitud de sistemas experimentales y diagnósticos, exploran el comportamiento del plasma midiendo magnitudes físicas, presión, temperatura, campos eléctricos, campos magnéticos, radiación emitida, partículas emitidas, etc. Estos sistemas de medida transforman las observaciones físicas en señales eléctricas, que se digitalizan y, las series temporales resultantes se transfieren y almacenan después de finalizar la descarga de operación.

En este trabajo se pretende comprobar la idoneidad y posibilidad de poder substituir sintéticamente una señal, en base al comportamiento anterior de las mismas y haciendo uso de la enorme cantidad de información disponible en la base de datos del dispositivo experimental de fusión nuclear.

Los principales objetivos del presente Trabajo Fin de Grado son los siguientes:

- **Estudiar y analizar** señales de evolución temporal en plasmas de fusión nuclear pertenecientes al dispositivo experimental TJ-II.
- **Parametrizar** señales y variables físicas mediante técnicas de regresión estadística.
- **Obtener** modelos resultantes similares en base a dicha parametrización.
- **Evaluar** la idoneidad de utilizar dichos modelos, para poder substituir predictivamente las señales, en caso de que un diagnóstico falle durante la adquisición de las mismas.

Regresión Paramétrica

Las técnicas de regresión paramétrica son aquellas que emplean distribuciones estadísticas con base en los datos que se tienen, presentando un resultado con cierto grado de fiabilidad [Pértega et al,2000], dado que deben de cumplir ciertos requisitos a la hora de aplicarle dichos modelos estadísticos.

A la hora de realizar una estimación acerca de las distintas ventajas posibles que presenta el empleo de dicha regresión [Cusiritati,2020], cabe destacar que, ante la presencia de cualquier variación en la información obtenida a lo largo del análisis de todas las descargas, el modelo se ve completamente afectado, ya que se constituyen dichas estructuras matemáticas a partir de los valores de secuencias previas.

Otro punto que destacar son los errores pequeños producidos tras la aplicación de dichas técnicas, dado que, al partir de un modelo matemático fundamentado en distribuciones estadísticas, estas se basan en aproximarse a la realidad lo máximo posible, sabiendo que dicha realidad siempre será más compleja, pero manteniendo gran similitud con el desarrollo de los hechos.

Entre las desventajas existentes por el empleo del uso de técnicas paramétricas, se deben de destacar dos principalmente. La primera de ellas es que los cálculos son difíciles de llevar cabo, dado que se necesita el uso de softwares bastante potentes, con el fin de llevar a cabo toda la comparación directa entre los valores presentes de descargas previas, y los distintos modelos estadísticos, estableciendo las distintas similitudes y adaptando todos los parámetros según se suman valores a la secuencia de descargas.

Por otro lado, mediante el empleo de dicha técnica, los datos que se pueden obtener son limitados, dado que se parte de una sucesión de valores previa, que al aplicarle un modelo estadístico que se aproxime a la forma que realmente queremos que tenga nuestro sistema, tendrá una estructura distinta, pero con valores limitados.

Otra característica que define dicho modelo es que todos sus algoritmos no están ligados al problema que se analiza, dado que, si en cualquier se quisiera realizar una modificación, el proceso no sería necesario volverlo a realizar desde el origen. A su vez, dicha ecuación final debe tener la posibilidad de poder ser evaluada en su nivel de complejidad.

Regresión Simbólica

Esta técnica inicialmente se propuso con el fin de realizar una aproximación empírica, la relación no lineal, existente y desconocida que presentan un conjunto de variables de entrada y una variable de salida [A.Martínez et al,2013].

La programación genética se ha empleado en varias ocasiones en un formato mixto, en el que se ha hecho uso de dicha técnica junto con modelos de corrección de errores. ARIMA es un modelo del que hicieron uso los científicos Lee y Tong en el año 2011, el cual se basaba en extraer la componente lineal de una serie de tiempo, a la vez que se empleaba dicha programación genética para obtener la componente no lineal.

La regresión simbólica, es una técnica comúnmente denominada como programación genética. Dicha tipología de análisis se emplea con la finalidad de buscar un patrón en una secuencia de datos, o averiguar la semejanza estadística existente entre una variable calificada como dependiente, y otras varias denominadas como independientes, de tal modo que como solución se encuentra buscar el modelo que define a la variable dependiente en función del resto de las variables [Moreno,2008].

Cabe destacar que actualmente, la programación genética es una modelo de la capacidad que se presenta a nivel computacional, presentando una evolución desmedida, donde lo que se consigue es que se lleve a cabo una tarea predefinida por un usuario. De este modo, dicho talento ingenieril, permite diseñar programas capaces de ejecutar procesos tras un proceso de inteligencia artificial, el cual se basa en la iteración de un proceso, hasta que sea capaz de encontrar el modelo que se ajuste más a realidad.

Otro aspecto a mencionar es que, entre los puntos más favorables de dicha técnica, se encuentra la resolución de problemas de alta complejidad, realizando una tarea que sería imposible de realizar por el ser humano.

También, es un proceso en el que se puede observar la secuencia de procesos uno a uno [Schmidt,2009], de tal modo que permite la posibilidad de análisis de evolución tras cada punto importante a analizar.

Otra característica que define dicho modelo es que todos sus algoritmos no están ligados al problema que se analiza, dado que, si en cualquier se quisiera realizar una modificación, el proceso no sería necesario volverlo a realizar desde el origen. A su vez, dicha ecuación final debe tener la posibilidad de poder ser evaluada en su nivel de complejidad.

Por otro lado, se ha de mencionar los puntos en contra que presenta dicho procedimiento matemático. El primero que se debe de destacar es que presenta soluciones con un rango finito. A su vez, existe una gran dificultad si se desean añadir restricciones al sistema, dado que la complejidad del sistema no permite realizar dicha tipología de modificaciones a la hora de realizar los ensayos.

En el caso de que se eligiera mal la ecuación que define el modelo matemático que se aproxima a la realidad, la resolución obtenida resultaría completamente errónea. Asimismo, lo que se pretende conseguir mediante el empleo de dicha técnica, es la obtención de un modelo matemático que se adapte en gran medida a la mayor cantidad de situaciones posibles, ya que el caso óptimo de un instante será muy difícil que lo sea en todos los instantes.

De este modo, una ecuación cualquiera que define el actual modelo se basa en el uso de sumas, restas, multiplicaciones, divisiones, senos, cosenos, ... por lo que es una técnica que no requiere de una gran dificultad a la hora de su aplicación, ya que las operaciones empleadas usualmente suelen ser fáciles de emplear, a pesar de que para obtener dicha ecuación final se requiera del empleo de softwares capaces de realizar innumerables cuentas por minuto.

$$Y = k_1 + k_2 * X_1 + k_3 * \sin(x_2) + \dots$$

Por último, hay que destacar que tanto el tiempo que se emplea a la hora de llevar a cabo una ejecución del proceso, como la robustez informática que presenta dicho método, hace que se prescinda de él en situaciones de bajo grado de dificultad, ya que resultaría un proceso inadecuado para la situación [Florida,2018].

Leyes de Potencia

Se define dicha técnica como la relación práctica entre dos valores, de tal modo que, si se aprecia una modificación en uno de los valores, el otro se verá afectado del mismo modo aplicándole un factor proporcional. De este modo, si se altera un valor, el otro sujeto se modifica como la potencia del primero.

Las leyes de potencia, o también llamadas de escala son un modelo matemático que se fundamentan exclusivamente en el producto de monomios, de tal modo que es una sucesión de elementos multiplicados entre sí, elevados a un determinado exponente, resultando del siguiente modo.

$$Y = k * X_1^{n_1} * X_2^{n_2} * X_3^{n_3} * X_4^{n_4} \dots$$

El ejemplo más básico que referencia las leyes de potencia con la realidad es aquel que tiene una única variable, y es el claro ejemplo de un cuadrado, ya que, si a dicha figura se le incrementa la longitud del lado de 3 a 6 metros, el área se verá afectada de tal modo que se transformaría de un valor de 9 m² a 36 m². Es decir, la ecuación resultaría del siguiente modo:

$$Y = k * X_1^{n_1}$$

Donde X e Y, serían las variables de interés. El símbolo α sería el exponente de la ley, y k una constante que afecta a todo el sistema por igual.

Además, se debe de destacar, que dicha técnica es útil cuando se emplea distribuciones en las que una cantidad pequeña de los valores se concentra en la zona superior de la gráfica, dominando así entorno al 95% de los recursos. Esto significaría que sucesos de altos valores son extraños, y por tanto un ocupan un pequeño tramo a lo largo de toda la gráfica [Dudley,2012].

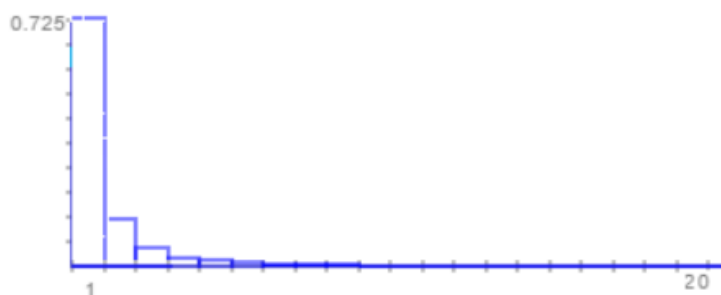


Ilustración 1. Distribución en leyes de potencia.

Entre los ejemplos reales más significativos a destacar se encuentran los siguientes:

- Repartición de los ingresos.
- Intensidad de un terremoto.
- Dimensión de localidades en función de sus habitantes.
- Repetición de términos en un mismo documento.

De este modo si se dibujan dos variables en escala logarítmica, y se observa una relación lineal, se llega a la conclusión de que presentan una estructura propia a las leyes de potencia.

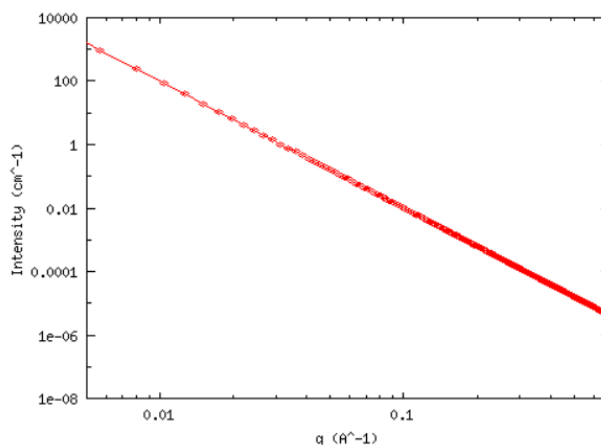


Ilustración 2. Gráfica Ordinary Least Square

Fuente: NIST.gov.

Leyes semejantes

Asimismo, existen distribuciones que presentan formas bastante similares a la distribución presente en las leyes de potencia. Este es el caso de la Ley de Pareto, ley no científica que organiza la densidad de los efectos de un hecho de tal modo que el 80% de estos, tienen como origen el 20% de las causas. Es decir, una gran cantidad de las actividades tienen poco impacto.

Otra ley similar, es la denominada como Ley de Zipf, la cual “representa la frecuencia de un evento en relación con su rango” [Glen,2016]. Es decir, lo que se pretende demostrar es que, si se eligiera un libro cualquiera, la palabra más empleada a lo largo de todo el libro aparecerá $N+2$ veces que la segunda palabra más usada. Y la segunda palabra más abundante se encontrará $M+2$ veces que la tercera, y así sucesivamente. Por lo que, si se analizara el caso más sencillo, dicha distribución sería la Ley de Potencia [Barker,1999].

Métodos analíticos

Ordinary Least Square

El presente método a desarrollar se caracteriza por ser un análisis estadístico, capaz de vincular la relación existente entre una o más variables independientes. La estimación se fundamenta en la minimización de la suma de los cuadrados con diferencia existente entre los resultados estimados y los reales [Lewis-Beck et al,1980].

Según define Michael Lewis-Beck, dado el supuesto caso de que la correspondencia entre ambas variables sea lineal, también calificada como regresión bivariable, se puede emplear una recta en una gráfica de dispersión. Dicha recta se caracterizaría por reflejar todos los puntos a lo largo de la evolución de la variable Y en función de la variable X, en los cuales el error estimado es 0 [Poston et al,2003].

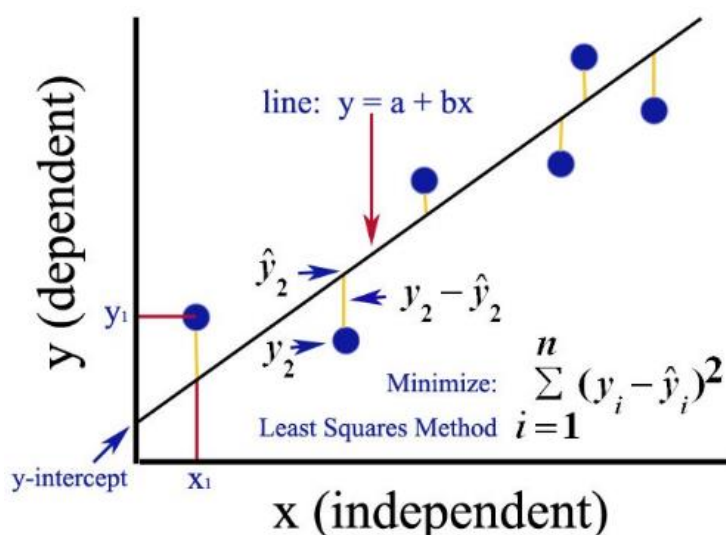


Ilustración 3. Grafica OLS

Fuente: Analytics Vidhya

Carl Friedrich Gauss realizó una primera suposición en la que mencionaba que, si se conocieran todos los valores de una variable Y exclusivamente, la predicción óptima sería la propia media de dichos valores [Gauss,1809]. No obstante, dicha suposición en una parte importante de los sucesos resultará errónea, difiriendo un el valor real del estimado en gran medida.

Sin embargo, si se obtuviera la información complementaria de una variable X , la duda que se plantea Lewis-Beck es, ¿cuánto mejor resultaría la predicción a la hora de minimizar los errores? La respuesta viene dada tras aplicar la técnica de la diferencia mediante mínimos cuadrados, dado que lo primero que se debe realizar es calcular dicha diferencia en los valores que toma la variable Y respecto de su valor estimado. Y una vez realizado dicho cálculo, se añaden los valores que toma la variable X , de tal modo que se calcula la nueva ecuación de regresión, conociendo de este modo la diferencia existente entre la suma total de cuadrados, con la suma de cuadrados explicada.

Ridge Regression

La técnica de Ridge Regression pretende realizar el análisis de los datos de regresión múltiple que padecen del efecto de multicolinealidad. En dichas situaciones, los cálculos mediante los mínimos cuadrados resultan inseguros, dado que presentan varianzas de gran valor, y por tanto es probable que se encuentre alejado del valor deseado. Sin embargo, si se aporta un grado de seguridad mayor, lo que ocurre es que dicha técnica reduce la aportación de errores estándar. De tal modo, la finalidad de dicha técnica es aportar estimaciones más fiables [NCSS,2012].

Es decir, un supuesto caso en el que una disolución formada por 3 compuestos distintos en las mismas proporciones. Al aplicarle los cálculos oportunos, haciendo uso de dicha técnica, si se produce alguna división entre 0, el procesamiento de datos debe ser abortado. Sin embargo, cuando los resultados obtenidos a lo largo de todas sus divisiones resultan distintos de 0, no se produce ninguna interrupción. Por lo que lo primero de todo es evaluar si se considera que existirá algún tipo de multicolinealidad entre las distintas variables.

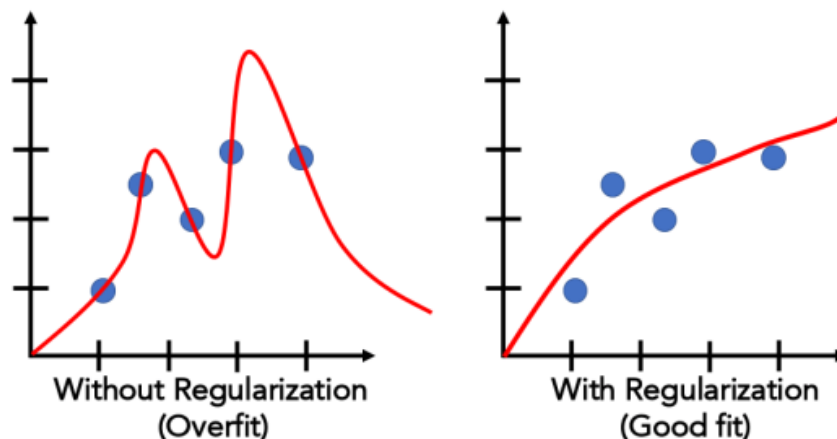


Ilustración 4. Gráfica Ridge Regression

El gran problema presente en dicha colinealidad resulta en que introduce estimaciones con poca precisión, agranda los errores estándar de los coeficientes de regresión, inflar los valores de las pruebas e incluso aportar valores no significativos, que el único efecto que causan es la degradación de la predictibilidad del modelo.

El modelo matemático con el cual se designa dicha técnica se escribe en formato de matriz, siendo esta:

$$Y = XB + e$$

Donde la variable Y resulta ser la variable dependiente, la X la variable independiente, B los distintos coeficientes de regresión que se deben de estimar, y la variable e son los errores en formato de residuos.

De tal modo, si partimos de que, a la hora de calcular los mínimos cuadrados ordinarios, los coeficientes se calculan mediante:

$$\hat{B} = (X'X)^{-1}X'Y$$

A partir del instante en el que se normaliza que, $X'X=R$, donde R representa la matriz de correlación de las variables independientes. Esto supone que,

$$E(\hat{B}) = B$$

y por tanto la matriz de covarianza se verá definida por la siguiente expresión,

$$V(\hat{B}) = \sigma^2 R^{-1}$$

donde se supone que el valor de $\sigma^2=1$. Por tanto, se llega a la conclusión de que el factor R al encontrarse en el denominador, a medida que se encuentra más cercano a 1, la varianza será mucho mayor.

Finalmente, si se añade un factor k , a la diagonal de la matriz de correlación, obtendríamos la siguiente expresión,

$$\tilde{B} = (R + kI)^{-1}X'Y$$

A raíz de dicha expresión, Hoerl y Kennard, creadores de dicho modelo matemático

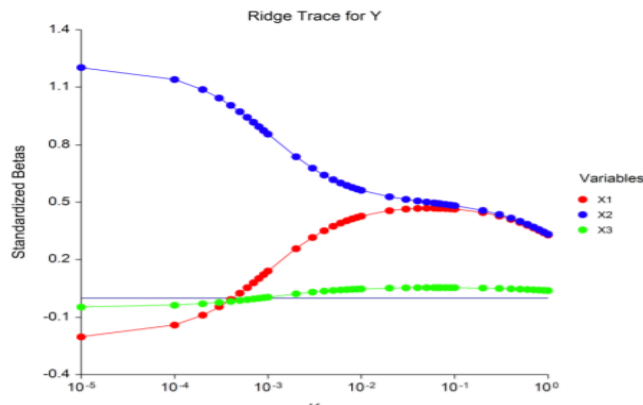


Ilustración 5. Factor k de Ridge Regression.

proporcionaron una gráfica mediante la cual permitía elegir el valor óptimo de k para la ejecución del ensayo [Hoerl et al,1970].

Total Least Square

La técnica TLS, presenta gran similitud con la técnica Ordinary Least Square. Ya que el diseño de ambos prácticamente presenta la misma base, a pesar de existir ciertas diferencias que permiten distinguir cuál es el método empleado en cada caso [Bannai,2018].

Total Least Square, es una técnica que se encarga de modificar de forma simétrica las matrices A y B. De tal modo que lo que se consigue obtener mediante dicho método es lo siguiente:

Minimizar el sistema: $\| [A \ B] - [\hat{A} \ \hat{B}]_F \|$, teniendo como sujeto $\hat{A}X = \hat{B}$.

De este modo, las condiciones presentes para que exista una solución exclusiva parte de que los valores dados en la matriz $[A \ B]$, son vectores singulares, que corresponden con los valores más pequeños [Golub et al,1970]. De este modo, se consigue que la forma óptima de para dicha aproximación es $[\hat{A} \ \hat{B}]$, la cual coincide con la norma de Frobenius en su forma de bajo rango no estructurado.

De este modo, retomando la comparativa entre el método OLS y TLS, la principal diferencia existente y visual a la hora de calcular la gráfica del modelo que se aproxima a la realidad, es que, en los puntos existentes en la gráfica, se trazan verticales en el método OLS, que en el método TLS se dibujan líneas ortogonales hacia la propia línea de regresión [Lawson et al,1987].

Un claro ejemplo práctico en el que se puede comprobar la diferencia existente entre una gráfica original, respecto de su gráfica aproximada mediante dicho método, sería la siguiente,

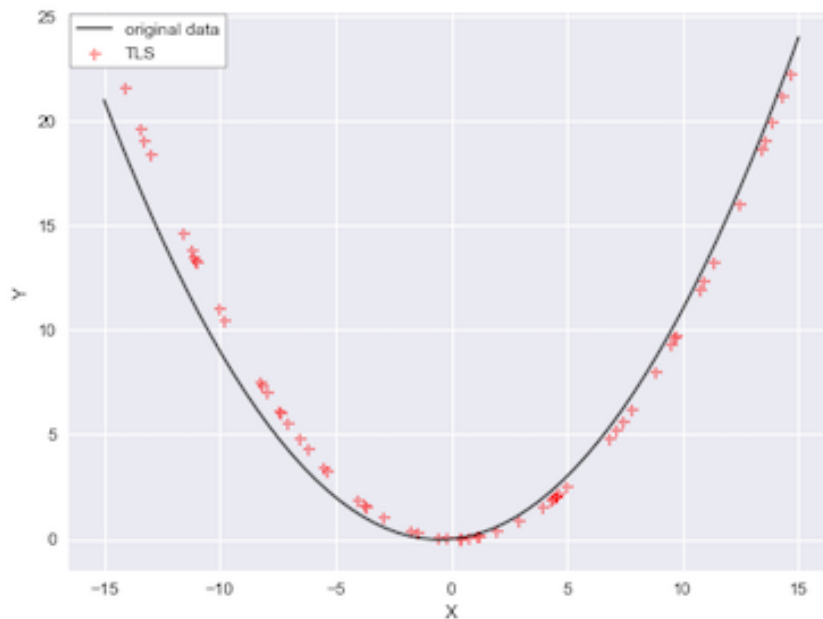


Ilustración 6. Grafica de TLS.

Regularized Total Least Square

En la actualidad, son considerados sistemas de error contaminados a aquellos sistemas que presentan un error tal que, $Ax = b$, donde A presenta un mal acondicionamiento. A su vez, se pueden resolver dichos sistemas mediante técnicas aproximadas a las de Tikhonov. En 1999, fue descubierto y mostrado a la sociedad un algoritmo capaz de hacer frente a la problemática referente a RTLS, el cual afectaba directamente al elemento multiplicador de Lagrange [Guo et al,2008].

La técnica empleada para llevar a cabo a la práctica se basa en el uso de la potencia inversa desplazada, la cual es una apreciación en base a la condición de limitación de regularización, sin necesidad de aplicar otro tipo de parámetros.

Además, la finalidad por la que se emplea dicha técnica, resultando óptimo frente a otras es porque tiene una gran capacidad para disminuir la cantidad de ruido en problemas de mínimos cuadrados. Sobre todo, el uso de esta técnica se considera útil cuando desconocemos la matriz de coeficientes.

Dicho desarrollo, se caracteriza por el truncamiento de las soluciones pequeñas presentes, con el fin de optimizar el modelo, de tal modo que al filtrar los valores se muestran en formato de descomposición de valor singular. Y finalmente, tras realizar dichas técnicas y aplicar procesos iterativos, se consiguen las soluciones del modelo TLS truncadas [Gratton et al,2013].

Robust Regression

Esta técnica presenta una utilidad en gran multitud de situaciones en la cual se esté empleando la regresión por métodos cuadrados. Su principal característica mediante la cual destaca dicho método es básicamente en no excluir aquellos valores que difieren en gran medida de los valores comunes entre los cuales se encuentra la sucesión estimada. Sin embargo, a pesar de ello lo que se realiza es una ponderación y reponderación mediante la cual se forma el modelo matemático [UCLA,2019].

En la actualidad, la técnica presente presenta distintas ramas, mediante las cuales se pueden llevar a cabo los cálculos de Robust Regression. En este aspecto podemos destacar la Regresión de Huber, de Ransac o de Theil Sen, todas ellas encargadas de mostrar todos los puntos que se reflejan en el registro de datos.

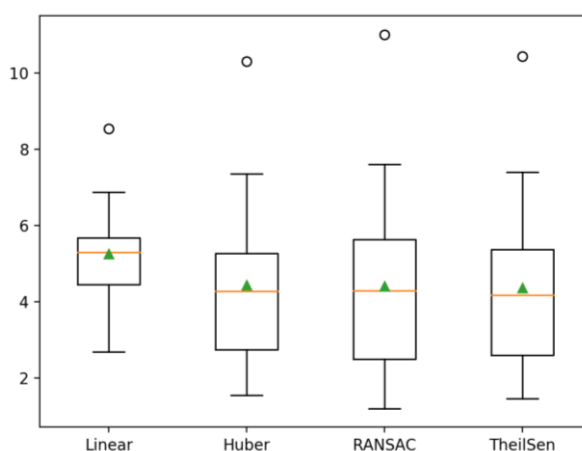


Ilustración 7. Gráfica de Robust Regression

Tras hacer referencia a la diferencia existente entre una técnica lineal y una técnica mediante de Robust Regression, este ejemplo muestra en un gráfico de cajas y bigotes, donde la caja supone el rango de valores que se suelen tomar en el espacio comprendido, y los bigotes la alta variación de valores poco comunes que se encuentran fuera del espacio común de tal modo, que en la técnica Robust Regression, todos los puntos se evalúan por igual.

De este modo, clarificando el diseño de las rectas que describen el modelo matemático óptimo para la ejecución de los valores, obteniendo en esos un menor error por mínimos cuadrados, se establecería del siguiente modo [Brownlee,2020].

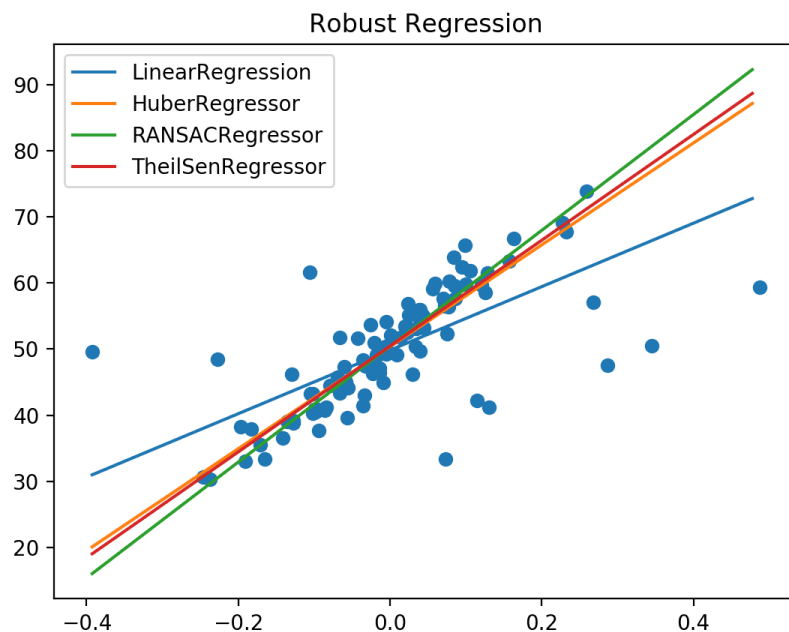


Ilustración 8.Rectas de distintos tipos de Robust Regression

Como conclusión, se observa que las rectas provenientes de las técnicas de Robust Regression, al tomar una mayor cantidad de valores, dibuja la línea a través de la mitad de todos ellos, mientras que la técnica lineal omite los valores más lejanos y se centra exclusivamente en el entorno de los valores más céntricos, es decir, una mayor delimitación de su rango.

Características y organización de la base de datos

Como punto de partida, se debe mencionar que el dispositivo TJ-II tras la realización de un ciclo de medidas donde se han parametrizado todas las señales captadas por los sensores, se obtiene un total de 494 archivos ASCII. Cada archivo ocupa un volumen de memoria del entorno de 30 a 40 megabytes, lo que implica una capacidad de 21 Gb de memoria en su totalidad.

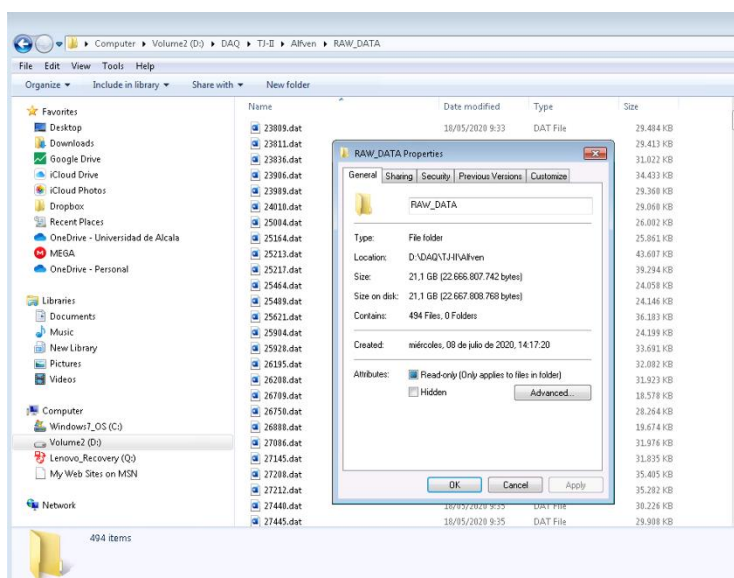


Ilustración 9. Memoria en uso por Base de datos

Una vez realizamos la descarga de los datos y trasparamos dichos valores a Matlab, se encuentra la estructura de datos brutos, la cual se define por presentar un total de 166000 muestras por cada una de las 8 señales que forman una única descarga.

Import - D:\DAQ\TJ-II\Alven\RAW_DATA\23809.dat

Column delimiters: Space Range: A2:H6602 Output Type: Replace unimportable cells with NaN

Variable Names Row: 1

Import Selection

	times	MBSCau	MBR1au	MBSP8au	Unlited T_eau	nI019m3	WpkI	IpkA	Halpba
	Number	Number	Number	Number	Number	Number	Number	Number	Number
1	Time (au)	MBSC (a.u.)	MBR1 (a.u.)	MBSP8 (a.u.)	T_e (0) (a.u.)	nI [10 ⁻¹ (19) m ⁻¹ (-3)]	Wp (kT)	Ip (kA)	Halpba
2	1019.999939	-0.1506715566	-0.02884004824	-0.2198249847	0.001376370317	-0.004547843710	-0.007480993401	0.004685061052	-0.007853699678
3	1020.000977	-0.1457875520	-0.07035408914	0.4714764655	0.001695359475	-0.004493224435	-0.007381699514	0.004490965046	-0.007495626342
4	1020.001953	-0.2043956071	-0.09477411211	0.2592088902	0.001943344600	-0.004420245066	-0.007286235224	0.004307726064	-0.007070353228
5	1020.002930	-0.1921185566	-0.07523809373	-0.2272002418	0.002128340632	-0.004329777323	-0.007188830059	0.004123649060	-0.006693921889
6	1020.003907	-0.1384615451	0.005347984943	-0.4758348763	0.002260545734	-0.004220116884	-0.007083225530	0.003926824313	-0.006171847694
7	1020.004944	-0.1409035474	0.01023199037	-0.2672086656	0.002331087831	-0.004109460860	-0.006981864106	0.003774010089	-0.005722667093
8	1020.005920	-0.1604395658	-0.03128205240	0.2182916552	0.00235913166	-0.003996716812	-0.006878602784	0.003551550560	-0.005263750907
9	1020.006958	-0.1579975635	-0.06791208684	0.3589810729	0.002339313272	-0.003880263539	-0.006766825449	0.003492054719	-0.004764701240
10	1020.007935	-0.1482295543	-0.05326007679	0.02600408532	0.002289806627	-0.003779096762	-0.006659697276	0.003156276885	-0.004283409591
11	1020.008972	-0.1555556612	-0.03868085929	-0.3441029787	0.002280895914	-0.003686309094	-0.006543837022	0.002948418725	-0.003759168554
12	1020.009949	-0.1335775405	0.01511599589	-0.3084896207	0.002112547401	-0.003618112067	-0.00642890426	0.002749778563	-0.003252881112
13	1020.010925	-0.115995124	-0.01907203905	0.59291664511	0.00200368870	-0.003575601367	-0.006320118438	0.002547925338	-0.002732862951
14	1020.011963	-0.1164853370	-0.04592406990	0.3506038487	0.001879582531	-0.003556226613	-0.006199308896	0.002328610521	-0.0021625200189
15	1020.012939	-0.1409035474	-0.09721611440	0.1538064033	0.001759717590	-0.003563458333	-0.006081807893	0.002120213583	-0.001824194323
16	1020.013977	-0.1335775405	-0.07768009603	-0.3601259887	0.001622433192	-0.003588146763	-0.005956064444	0.001893228618	-0.001072976505
17	1020.014954	-0.115995124	0.01023199037	-0.3342881501	0.001476953155	-0.003611995608	-0.005835890304	0.001675054431	-0.000670187151
18	1020.015930	-0.09206349403	0.01511599589	0.1535301059	0.001308972365	-0.003652871819	-0.005713978828	0.001452180324	-0.0002255385043
19	1020.016966	-0.09594749862	-0.03616605699	0.3917956352	0.001098463661	-0.003681161674	-0.005582488608	0.001209902010	5.665820208e-05
20	1020.017944	-0.1115995124	-0.09721611440	0.1601981670	0.0008632408316	-0.003694060799	-0.005456968676	0.0009764196384	0.0001739437375

Ilustración 10. Inicio de base de datos en Matlab

	timesms	MIRSau	MBRSau	MEPSau	T_eau	n1019m3	WpkJ	IpkA	Halpha
166783	1186.781006	-0.4168498218	0.05563003772	0.1450158954	-0.03959609941	1.039573789	0.1229387373	-0.008032478392	0.04574034363
166784	1186.781982	-0.4046398103	0.1176861026	-0.1860551983	-0.03961459920	1.039289474	0.1227895543	-0.007947967558	0.04560532795
166785	1186.782959	-0.3948718011	0.09326007962	-0.273502141	-0.03960794897	1.039106607	0.1226396188	-0.007855478789	0.04551422133
166786	1186.783936	-0.4192918241	0.0246040601	-0.06399211679	-0.03959216603	1.039001763	0.1224898532	-0.007849726185	0.045461742598
166787	1186.784912	-0.4339438379	0.0199999955	0.3008664548	-0.03954511741	1.038987160	0.1223375723	-0.007759584032	0.04546523362
166788	1186.785889	-0.4192918241	0.08349206299	0.5147706270	-0.03950512782	1.039021492	0.1221855059	-0.0076757926618	0.04546787927
166789	1186.786867	-0.4021978080	0.1127960980	0.3709360361	-0.03946483508	1.039111137	0.1220136285	-0.006831584033	0.05059882795
166790	1186.787844	-0.4168498218	0.04930403084	0.06949009746	-0.03944199905	1.039219975	0.1218601614	-0.006632928271	0.05239731073
166791	1186.788820	-0.4412698440	-0.006962026639	-0.2016009539	-0.03943929449	1.039341331	0.1217060611	-0.006438613404	0.05418404564
166792	1186.789797	-0.4339438379	0.0517460313	-0.2682614625	-0.03946432099	1.039460421	0.1215513372	-0.006250512786	0.05599930882
166793	1186.790774	-0.4144078195	0.1225641072	0.04056917092	-0.03953108640	1.039573789	0.1213766187	-0.006048641657	0.05800123140
166794	1186.791752	-0.4192918241	0.07616601510	0.4396601915	-0.0396423761	1.039643288	0.1212207004	-0.00587997252	0.05947894937
166795	1186.792729	-0.4339438379	0.01023199037	0.5110434890	-0.03979993239	1.039684534	0.1210642308	-0.005721892696	0.06121146679
166796	1186.793706	-0.4510378140	0.02488400601	0.2736392021	-0.03999542817	1.039702216	0.1209072471	-0.005572012160	0.06257651001
166797	1186.794683	-0.4388278425	0.08593407273	-0.07981520146	-0.04022292420	1.039701462	0.1207497567	-0.005427618045	0.06375461072
166798	1186.795660	-0.4241758287	0.09326007962	-0.2652247250	-0.04047459246	1.039686799	0.1205917895	-0.005285991821	0.06472632389
166799	1186.796637	-0.4363858402	0.02732600830	-0.1019218042	-0.04077886742	1.039659739	0.1204335269	-0.005126586184	0.06554836151
166800	1186.797614	-0.460895631	-0.006962026639	0.2822222114	-0.04105394706	1.039631405	0.1202345896	-0.00498116281	0.06601708382
166801	1186.798591	-0.4587638608	0.04866202854	0.4956939816	-0.04133076593	1.039604763	0.1200952404	-0.004831119675	0.06621840596
166802	1186.799567	-0.4388278425	0.1103540957	0.4127719104	-0.04159951583	1.039584041	0.1199355274	-0.004672378767	0.06613308191

Ilustración 11. Fin de base de datos en Matlab.

Preparación y Preprocesamiento de las señales

El procedimiento que se lleva a cabo una vez se obtiene la base de datos en bruto, se compone de una secuencia de procesos la cual requiere de la ejecución de una serie de comandos en Matlab, ya que sin dicho programa sería imposible operar con los valores en bruto.

El primer paso que se debe de llevar a cabo es la aplicación de la transformada de Wavelet – Haar [Pereira,2010] la cual es capaz de descomponer en elementos de tiempo-frecuencia de modo casi simultáneo. Además, la transformada de Haar se considera una variante simple dentro del análisis wavelet discreto, el cual permite disminuir el número de valores brutos pertenecientes a una señal discreta. Este procedimiento se basa en el cálculo de la media en agrupaciones de 2 elementos, por lo que la base de datos debe de ser una potencia de 2 necesariamente. Dicha transformada realiza una doble descomposición formando así dos sub-señales, donde cada una de estas se define como coeficientes de aproximación y coeficientes de detalle.

$$\begin{aligned}
 [CL] &= \text{wavedec}(Te, \text{nivel}, 'haar'); \\
 \text{cproxima4} &= \text{appcoef}(C, L, 'haar'); \\
 \text{cdetalle4} &= \text{detcoef}(C, L, \text{nivel});
 \end{aligned}$$

De este modo, cabe destacar que en el primer nivel de dicha función encontramos $N/2$ wavelets de Haar, expresadas del siguiente modo:

$$\begin{aligned}
 w_1^1 &:= \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0, \dots, 0 \right) \\
 w_2^1 &:= \left(0, 0, \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0, \dots, 0 \right) \\
 &\vdots \\
 w_{N/2}^1 &:= \left(0, \dots, 0, \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \right).
 \end{aligned}$$

Al mismo tiempo, se obtiene como conclusión que la agrupación de vectores constituyen un sistema ortogonal. Partiendo de la base de que cada uno de ellos son ortogonales.

$$\begin{aligned}
 w_1^1 \cdot w_2^1 &= \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0, \dots, 0\right) \cdot \left(0, 0, \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0, \dots, 0\right) = \frac{1}{\sqrt{2}} \cdot 0 + \frac{-1}{\sqrt{2}} \cdot 0 + \dots + 0 = 0 \\
 w_1^1 \cdot w_3^1 &= \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0, \dots, 0\right) \cdot \left(0, 0, 0, 0, \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0, \dots, 0\right) = \frac{1}{\sqrt{2}} \cdot 0 + \frac{-1}{\sqrt{2}} \cdot 0 + \dots + 0 = 0 \\
 &\vdots \\
 w_{N/2-1}^1 \cdot w_{N/2}^1 &= \left(0, \dots, 0, \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0, 0\right) \cdot \left(0, \dots, 0, \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right) = 0 + \dots + \frac{1}{\sqrt{2}} \cdot 0 + \frac{-1}{\sqrt{2}} \cdot 0 = 0.
 \end{aligned}$$

No obstante, al calcular la norma de cada uno de los vectores que forman dicho sistema, el resultado es 1 [Olivera,2017].

$$\|w_j^1\| = \sqrt{\left(\frac{1}{\sqrt{2}}\right)^2 + \left(\frac{-1}{\sqrt{2}}\right)^2 + 0^2 + \dots + 0^2} = \sqrt{\frac{1}{2} + \frac{1}{2}} = \sqrt{1} = 1$$

Por lo tanto, tras conocer la teoría que define el funcionamiento interno de dicha transformada, la base total de datos se reduce a un valor de 128 muestras. Esto implica que como se ha mencionado previamente, se conserve la forma estructural de la señal, y almacenando los coeficientes de aproximación. Finalmente, se agrupan todas las señales y descargas en un único archivo Matlab, el cual ocupa una memoria de 3,5 Megabytes y una estructura con mayor facilidad para su manejo y uso a la hora de realizar cálculos y tratar con ellos durante su análisis.

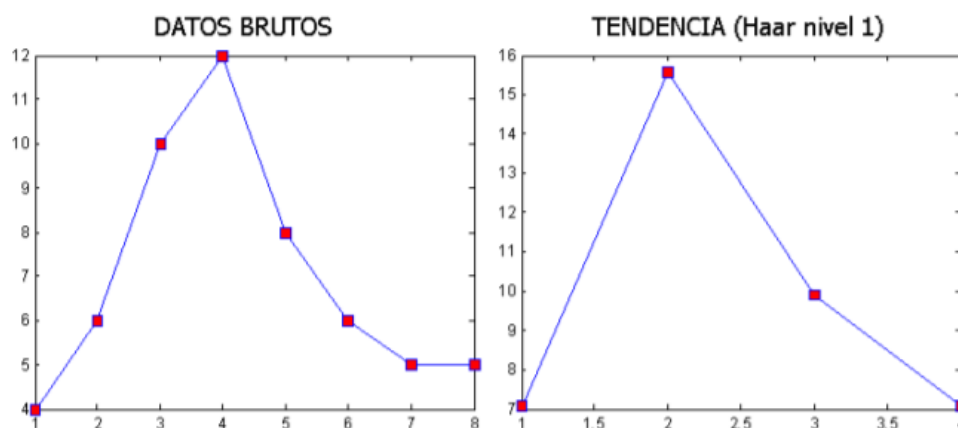


Ilustración 12. Comparativa Transformada de Haar

Una vez se ha aplicado las transformadas mencionadas previamente con el fin de reducir el tamaño de la base de datos, se procede a realizar la preparación de los datos de tal modo, que

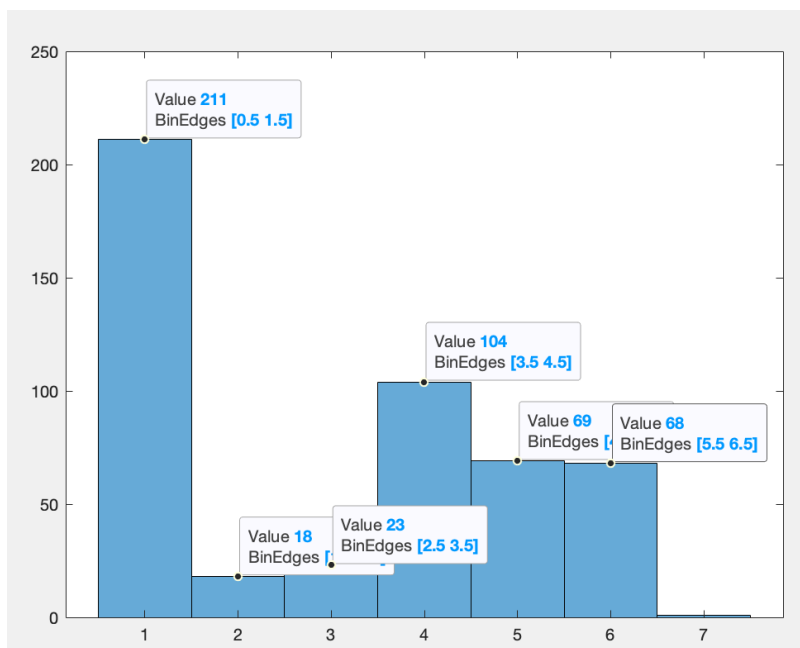


Ilustración 13. Descargas por tipo de Calentamiento

los valores al presentar valores tanto positivos como negativos, resulta imposible aplicar expresiones como el logaritmo, ya que para su aplicación los valores negativos carecen de sentido.

De este modo, se realiza una normalización de los datos mediante el método de máximo y mínimo específico, el cual se basa en relacionar todos los valores que presenta una base de datos, procediendo a que estos tomen cifras comprendidas entre 1 y 10. De este modo, se podrá aplicar la función logaritmo en base 10 de cualquier valor de la base de datos normalizada, ya que como resultado final se obtendrá un valor comprendido entre 0 y 1.

Por último, la base de datos que se ha empleado y con la cual se va a proceder a trabajar se analiza con el fin de poder conocer estos valores en función del tipo de calentamiento que se le aplica, y ser conscientes cuantas descargas existen de cada tipo, de este modo resultaría que el tipo de calentamiento 1 presenta doscientas diecisiete (muestras), del tipo 2 se tienen doce (12) muestras, del calentamiento 3 se presentan veinticuatro (24), del tipo 4 se muestrean ciento ocho (108), del tipo 5 sesenta (60) muestras, del calentamiento 6 se disponen setenta y una (71) y del calentamiento 7 exclusivamente una sola muestra.

Metodología de Procesos

La evaluación de las técnicas de regresión simbólica y leyes de potencia, presentan un patrón común mediante el cual se rigen ambos procedimientos. En ambos casos, se hace uso de una única base de datos, la cual permite ser la base comparativa, y esto implica que se le deben de aplicar procedimientos comunes.

La base de datos formada por las 494 descargas es un documento en formato struct llamado “discharge_raw”, en el están recogidas todas las variables en matrices internas de cada una de las descargas que se reflejan en el eje Y del struct, además están divididas internamente en 7 grupos en función del tipo de calentamiento que se le ha aplicado antes de obtener los valores presentes.

discharge_raw							
1x494 struct with 7 fields							
Fields	number	Te	n	Wp	Ip	Halp	Calentamiento
1	23809	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
2	23811	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
3	23836	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
4	23906	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
5	23989	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
6	24010	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
7	25004	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	3
8	25164	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	2
9	25213	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
10	25217	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
11	25464	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
12	25489	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
13	25621	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
14	25904	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	3
15	25928	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	3
16	26195	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	3
17	26208	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	3
18	26709	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
19	26750	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
20	26888	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
21	27086	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
22	27145	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
23	27208	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
24	27212	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1
25	27440	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	128x1 doub...	1

Ilustración 14. Listado de descargas en base de datos normalizada

De este modo, se procede a realizar un análisis de cada una de las 5 variables a analizar por descarga, estas son temperatura, densidad, energía diamagnética, corriente del plasma y H-alpha. Por lo que, cuando se quiera realizar el análisis de la forma que tomaría la curva de dicha variable en una descarga específica, se debería de hacer uso para cada una de ellas de la función, “find previous discharges more similar to.m”.

```

1 function resultado = find_previous_discharges_more_similar_to(discharge);
2 clc;
3 load('discharge_raw.mat');
4
5 for i=1:length(discharge_raw);
6     shot(i,1) = discharge_raw(i).number;
7     shot(i,2) = discharge_raw(i).Calentamiento;
8 end;
9
10 indi = find(shot==discharge);
11
12 for h=1:indi-1;
13     simi_Te(h) = pdist2(discharge_raw(indi).Te',discharge_raw(h).Te','cosine');
14     simi_n(h) = pdist2(discharge_raw(indi).n',discharge_raw(h).n','cosine');
15     simi_Wp(h) = pdist2(discharge_raw(indi).Wp',discharge_raw(h).Wp','cosine');
16     simi_Ip(h) = pdist2(discharge_raw(indi).Ip',discharge_raw(h).Ip','cosine');
17     simi_Halphi(h) = pdist2(discharge_raw(indi).Halphi',discharge_raw(h).Halphi','cosine');
18 end;
19 [m_Te, i_Te] = sort(simi_Te);
20 [m_n, i_n] = sort(simi_n);
21 [m_Wp, i_Wp] = sort(simi_Wp);
22 [m_Ip, i_Ip] = sort(simi_Ip);
23 [m_Halphi, i_Halphi] = sort(simi_Halphi);
24
25 resultado(1:5,:) = [shot(i_Te(1:5)',1) i_Te(1:5)' m_Te(1:5)' shot(i_Te(1:5)',2)];
26 resultado(6:10,:) = [shot(i_n(1:5)',1) i_n(1:5)' m_n(1:5)' shot(i_n(1:5)',2)];
27 resultado(11:15,:) = [shot(i_Wp(1:5)',1) i_Wp(1:5)' m_Wp(1:5)' shot(i_Wp(1:5)',2)];
28 resultado(16:20,:) = [shot(i_Ip(1:5)',1) i_Ip(1:5)' m_Ip(1:5)' shot(i_Ip(1:5)',2)];
29 resultado(21:25,:) = [shot(i_Halphi(1:5)',1) i_Halphi(1:5)' m_Halphi(1:5)' shot(i_Halphi(1:5)',2)];

```

Ilustración 15.Funcion find previous discharges.m

El uso de esta función implica que cuando se pretenda analizar la variable de la temperatura en una descarga de tipo de calentamiento 1, dicha función será la encargada de buscar en la base de descargas previas a la que se solicita, exclusivamente en la variable de temperatura, y se establecerán las 5 descargas más similares de cada una las 5 variables a analizar, lo único que se omitirá la variable que se desea buscar, de este modo en base a las otras 4 variables se selecciona la descarga más similar.

En la ilustración 15, para el tipo de calentamiento 1 se ha elegido la descarga 33308, y si se quiere analizar la variable temperatura, se crearía un modelo matemático a partir de la variable de densidad de la descarga 33282, la variable energía diamagnética de la descarga 33263, la variable corriente del plasma de la descarga 33290 y la variable Ha de la descarga 33282. De este modo, se podría estimar un modelo matemático que posteriormente se comparará con la realidad para corroborar si es aproximado, o si presenta un error elevado.

Signals and discharges most similar to 33308 - Index 109 Type 1:

Signal	Discharge	Index	Similarity	Type
Te	33273	103	0.001344	1
Te	33263	102	0.001410	1
Te	33281	104	0.001604	1
Te	33290	108	0.001981	1
Te	33282	105	0.002340	1
n	33282	105	0.002545	1
n	33263	102	0.003055	1
n	33281	104	0.003587	1
n	33290	108	0.004002	1
n	33239	100	0.004394	1
Wp	33263	102	0.004274	1
Wp	33281	104	0.004499	1
Wp	33282	105	0.005310	1
Wp	33290	108	0.005750	1
Wp	33273	103	0.006206	1
Ip	33290	108	0.020609	1
Ip	33239	100	0.021785	1
Ip	33209	97	0.023650	1
Ip	33244	101	0.024165	1
Ip	33222	98	0.031346	1
Ha	33282	105	0.001140	1
Ha	33290	108	0.001449	1
Ha	33281	104	0.002689	1
Ha	33288	107	0.003135	1
Ha	33263	102	0.003644	1

Ilustración 16.Resultado de descargas similares

Desarrollo Regresión Simbólica

Tras la explicación desarrollada en el anterior apartado, y una vez teniendo las descargas a partir de las cuales se fundamenta en modelo matemático a desarrollar, la técnica de regresión simbólica se ejecutará mediante el uso de dos softwares diferentes, uno será el Matlab, y el otro será la versión Nutonian Eureka desktop versión 1.24.0.9367 bajo una licencia de prueba [Smidt, et al., 2013].

Matlab será la herramienta fundamental, ya que además de emplearlo para la obtención de las descargas del apartado anterior, también nos permitirá ejecutar el modelo matemático diseñado mediante el otro software y realizar una gráfica comparativa en la que podamos mostrar el error existente entre la curva real y la procedente del modelo matemático.

Por otra parte, Eureka es un software con un potencial matemático computacional de alto calibre, dado que es el programa informático capaz de ejecutar miles de operaciones por minuto, de este modo permite realizar ejecuciones fundamentadas en el ensayo prueba-error, de tal forma que se consigue un modelo matemático que a medida que transcurre el tiempo se eleva la complejidad del sistema, y como consecuencia la precisión de este.

Por lo tanto, el proceso llevado para obtener la gráfica que se está desarrollando a lo largo de todo el documento, el análisis de la variable temperatura, en la descarga 33308 con tipo de calentamiento 1. Partiendo de la ilustración 15, se procede a implementar los valores internos de cada variable,

discharge_raw(102).Wp		discharge_raw(105).n		discharge_raw(108).Ip		discharge_raw(105).Halp	
1	2	1	2	1	2	1	2
1	1.6280	1	1.0375	1	4.8175	1	1.0070
2	1.3110	2	1.0545	2	4.8329	2	1
3	1	3	1	3	4.8285	3	1.0170
4	1.1529	4	1.0663	4	4.8261	4	1.0226
5	1.0578	5	1.1300	5	4.6921	5	1.0703
6	1.6280	6	2.9392	6	4.4477	6	3.7537
7	3.6437	7	6.6585	7	4.2332	7	5.0128
8	6.7947	8	7.8455	8	4.1964	8	4.8203
9	6.5921	9	7.6783	9	3.7804	9	8.1729
10	5.6596	10	6.8280	10	3.5953	10	9.3699
11	5.3499	11	5.7669	11	3.3425	11	8.7266
12	4.9021	12	5.1936	12	3.1821	12	7.8410
13	4.4196	13	5.0882	13	3.2212	13	7.2862
14	4.6944	14	5.3769	14	3.7557	14	7.1287
15	5.2767	15	6.1053	15	4.7054	15	7.2775
16	5.3785	16	6.7062	16	5.2390	16	8.2451
17	5.6405	17	7.0816	17	5.2877	17	9.0665
18	6.3952	18	7.2249	18	5.3268	18	9.5184
19	7.1036	19	7.2509	19	5.3263	19	9.5682
20	7.0001	20	7.2077	20	5.4826	20	9.9029
21	6.9731	21	7.0802	21	5.7633	21	10
22	7.5531	22	6.9709	22	5.6788	22	9.7475
23	6.8958	23	6.4458	23	5.3961	23	9.5728

Ilustración 17. Valores de cada variable con su respectiva descarga

Una vez se toman los 128 valores que constituyen cada una de las variables que constituyen estas columnas, se insertan dichos valores en el programa Eureka, es decir, cada columna que constituye una variable en el programa estará formada por 4 veces, debido al número de variables a analizar, y a sus 128 valores asociados a cada una de ellas, por lo que cada columna presentaría 512 filas con valores, salvo que algún número de descarga fuera repetido y por tanto no sería necesario implementarlo, aunque si se añadiera no existiría ningún efecto adverso.

	A	B	C	D	E
info	Temperatura	Densidad	Energía diamagnética	Corriente del Plasma	H alpha
name	T_e	n	W_p	I_p	H_a
1		1.0374682197...	1.1200187288...	3.2533680264...	1.0069680379...
2		1.0545290061...	1.0008109439...	3.2393947411...	1
3		1	1.1268967020...	3.2484894012...	1.0170095438...
4		1.0663093289...	1	3.2678348810...	1.0226357638...
5		1.1299767048...	1.2083452024...	3.2728419435...	1.0702592998...
6		2.9391539187...	1.2751006355...	3.2889473102...	3.7536950551...
7		6.6584594385...	2.1375690657...	3.8414679392...	5.0127900417...
8		7.8454967782...	3.7913225230...	4.6985577345...	4.8203481784...
9		7.6782569254...	5.3740335512...	5.8594587830...	8.1728582907...
10		6.8280184720...	5.8372850433...	6.0611145012...	9.3699148186...
11		5.7668825385...	5.5169350603...	5.9216181332...	8.7265998464...
12		5.1935633033...	5.0785043799...	6.2130615132...	7.8410314628...
13		5.0882057883...	4.7944918095...	6.4619841881...	7.2862073786...
14		5.3768501717...	4.7525844038...	6.9967264710...	7.1286514582...
15		6.1052642730...	5.2259872625...	7.6294333120...	7.2775083605...
16		6.7062456782...	5.9234722456...	7.5522867022...	8.2450666893...
17		7.0815824236...	6.4436691945...	7.0995459573...	9.0664850029...
18		7.2249167261...	6.9596988933...	6.6680351230...	9.5183801243...
19		7.2508623201...	7.1008756358...	6.5238205987...	9.5682174131...
20		7.2076847254...	7.1435599424...	6.0981017996...	9.9028846298...
21		7.0802272432...	7.4258295814...	5.6973335931...	10
22		6.9708747017...	7.0756797746...	5.0433925682...	9.7475235089...
23		6.8157670953...	6.5963396616...	4.6525581380...	9.5728434859...
24		6.7097176630...	6.2832121077...	4.6304405638...	9.2945350221...
25		6.6128575540...	6.3354337681...	4.0972137711...	9.115681308225...
26		6.4839848526...	6.0452346485...	3.2479291484...	8.9521485823...
27		6.3644823543...	5.7191333422...	2.7279319323...	8.6605969856...
28		6.2294248335...	5.7350236156...	2.5664418564...	8.6222616650...
29		6.0707485486...	5.7276652368...	2.1080879621...	8.5007328463...
30		5.9889083065...	5.5592688955...	1.8139577834...	8.311734119127...
31		5.8842748446...	5.7260086338...	1.8538810704...	7.6854281445...
32		6.1763123537...	5.7165333267...	2.0390265812...	5.7480059112...
33		6.5140378774...	5.6275736950...	1.9062686682...	4.8877166993...
34		6.7841455851...	5.2896091218...	2.1599670938...	4.8603549777...
35		6.8430271772...	5.1306016928...	2.1591234505...	5.1630424251...
36		6.8296875339...	4.9103548745...	1.4792920920...	5.3055792442...
37		6.7771529610...	4.5674057362...	1.0500831465...	5.2689262550...
38		6.7278720285...	4.3793533738...	1	5.1529164729...
39		6.6612219979...	4.7874290708...	1.2964136041...	5.1602499207...
40		6.5786666447...	4.7361075895...	1.3789991475...	5.2323171514...
41		6.5564771047...	4.4805617481...	1.7820193109...	5.1642434276...
42		6.4397276760...	4.3863648905...	1.675361138408...	5.1741341829...
43		6.4591528960...	4.8859526002...	1.7858050641...	5.1160285122...
44		6.4673605732...	4.8783961124...	1.7239410944...	5.1593385287...
45		6.5135576995...	4.7552590215...	1.7151202867...	5.2128968360...
46		6.5436179800...	4.8638295014...	1.4974036558...	5.1567909723...
47		6.4995785653...	4.9489713708...	1.3469776232...	5.2217118585...
48		6.4743616379...	4.9819980718...	1.5231766661...	5.2324960800...
49		6.5292684790...	5.2727085466...	1.7163480030...	5.1983140239...
50		6.5586934216...	5.6121995226...	2.0670987006...	5.2121606665...
51		6.5923794930...	5.8840783361...	2.2659553498...	5.3218101139...
52		6.6285071936...	5.9619553677...	2.3121611388...	5.2725783619...
53		6.6319292655...	6.2562669780...	2.5609013983...	5.3976784887...
54		6.6294272868...	6.4088934206...	2.8641184954...	5.3532518782...
55		6.6200023544...	6.0024876407...	2.5555200442...	5.4084045599...

Ilustración 18. Base de datos en Eureka

A continuación, se define la función que se quiere conocer en función del resto de las variables, en este caso será la temperatura en función del resto de las variables. Además, el software permite elegir las funciones de las que se va a hacer uso para llegar al modelo matemático final. Para el desarrollo de dicho análisis, se ha hecho uso de constantes, variables de entrada, sumas, restas, multiplicaciones, divisiones, senos y cosenos, de tal modo que resulta un sistema con un nivel de complejidad elevado, sin llegar a tener un grado tan elevado como para que al trasladar la ecuación a otro software de algún tipo de error.

The Target Expression:

Search for a formula $f()$ that satisfies the equation:
[See Examples](#)

$Te = f(n, Wp, Ip, Ha)$

Primary Options:

Formula building-blocks:	Name	Complexity
Basic		
<input checked="" type="checkbox"/>	Constant	1
<input type="checkbox"/>	Integer Constant	1
<input checked="" type="checkbox"/>	Input Variable	1
<input checked="" type="checkbox"/>	Addition	1
<input checked="" type="checkbox"/>	Subtraction	1
<input checked="" type="checkbox"/>	Multiplication	1
<input checked="" type="checkbox"/>	Division	2
<input type="checkbox"/>	Negation	1
Trigonometry		
<input checked="" type="checkbox"/>	Sine	3
<input checked="" type="checkbox"/>	Cosine	3
<input type="checkbox"/>	Tangent	4

Ilustración 19. Definición de variables dependiente e independientes al definir el modelo

Posteriormente, al entrar en la ventana “Start Search”, se presiona sobre el botón “Run”, y la ejecución interna de operaciones se inicia, de tal modo que a la derecha se refleja cómo se ajusta el sistema en función de la complejidad y del error estimado que se espera, lo que implica largos períodos de tiempo por ejecución ya que cada ciclo de variable se puede demorar en torno a los 35 ó 45 minutos, lo que hace que sea un proceso lento pero eficaz.

Run the formula search: Run Pause Stop

Progress & performance

Time	
Search time	36m 18s
CPU cores	0
Performance	
Generations	489190
Generations/sec	0
Formula evaluations	3.6e10
Evaluations/sec	0.00 M
Confidence	
Stability	58.1%
Maturity	71.5%
Percent converged	100%

Ilustración 20. Tiempo de duración en obtener el modelo

En el caso de la ilustración 19, el tiempo que ha llevado generar el modelo matemático de dicha variable de temperatura en función del resto es de 36 minutos y 18 segundos. Además, se puede comprobar que se han realizado 489190 cuentas a lo largo de todo el período de tiempo, con una cantidad de 3.6e10 factores empleados.

Por otra parte, se refleja la seguridad que proporciona dicho modelo matemático, ya que el informe muestra una estabilidad del 58,1% y una denominada madurez, que realmente refleja lo estable que es la señal en la totalidad del período estudiado del 71,5%.

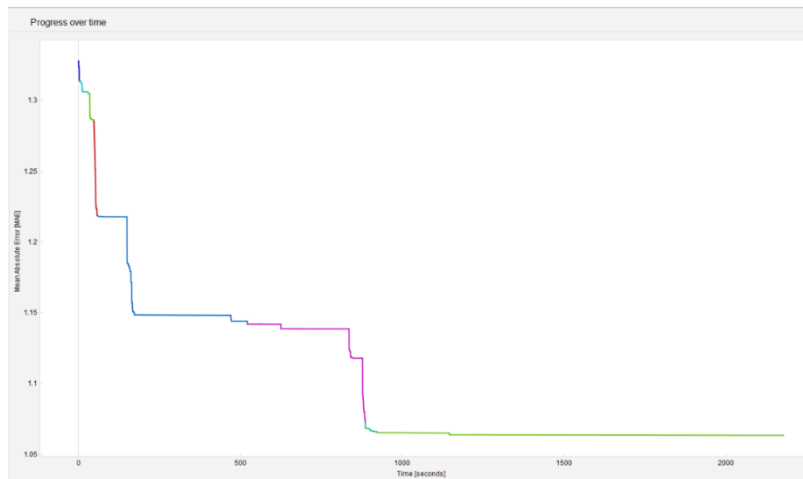


Ilustración 21. Curva error vs Tiempo

A continuación, en la pestaña “Results”, se pueden apreciar las distintas ecuaciones que generan distintos modelos matemáticos que se adaptan al sistema generado, de tal modo que se puede elegir la ecuación en función de la complejidad y exactitud que se desee.

Size	Fit	Solution
29	0.240	$Te = 4.77 Wp + 1.13 Ha + 0.0378 Wp^3 - 4.46 - 0.102 nHa - 0.757 Wp^2$
27	0.244	$Te = Ha + 4.76 Wp + 0.0373 Wp^3 - 4.27 - 0.0858 nHa - 0.755 Wp^2$
32	0.238	$Te = 4.1 Wp + 1.08 Ha - 4.21 - 0.0969 nHa - 0.33 Wp^2 - 3.54 \sin(0.0349 Wp^2)$
41	0.236	$Te = 5.01 Wp + 0.723 Ha + 0.0386 Wp^3 + 0.00058 WpHan^2 - 4.36 - 0.801 Wp^2 - 0.01 Han^2$
18	0.291	$Te = 5.29 n + 0.368 Ha + 28.5 \cos(0.219 n) - 32.7$
23	0.287	$Te = 5.21 n + 0.386 Ha + 0.0365 n^3 - 4 - 0.811 n^2$
25	0.284	$Te = 6.21 n + 0.0446 n^3 + 0.0255 Ha^2 - 4.54 - 0.971 n^2$
16	0.305	$Te = 5.93 + 1.33 Ha + \frac{-6.43}{n} - 0.13 nHa$
13	0.323	$Te = n + 2.36 Ha - 2.1 - 0.276 nHa$
10	0.466	$Te = 4.41 + 0.41 Ha + \cos(n)$
9	0.588	$Te = 9.47 + 0.381 Ha - 0.615 n$
8	0.626	$Te = 6.85 + 0.87 \cos(n)$
6	0.640	$Te = 6.88 + \cos(n)$
5	0.738	$Te = 11.3 - 0.576 n$
3	0.949	$Te = 0.669 + Ha$
1	1.000	$Te = 6.7$

Ilustración 22. Ecuaciones del modelo

Por lo que, si se eligiera la ecuación valorada como 29, es la que el software reconoce que, en función de la complejidad y el esfuerzo empleado para desarrollarla, a la vez que se refleja una gran exactitud del modelo con la supuesta curva deseada, lo cual queda reflejado en la ilustración 22.

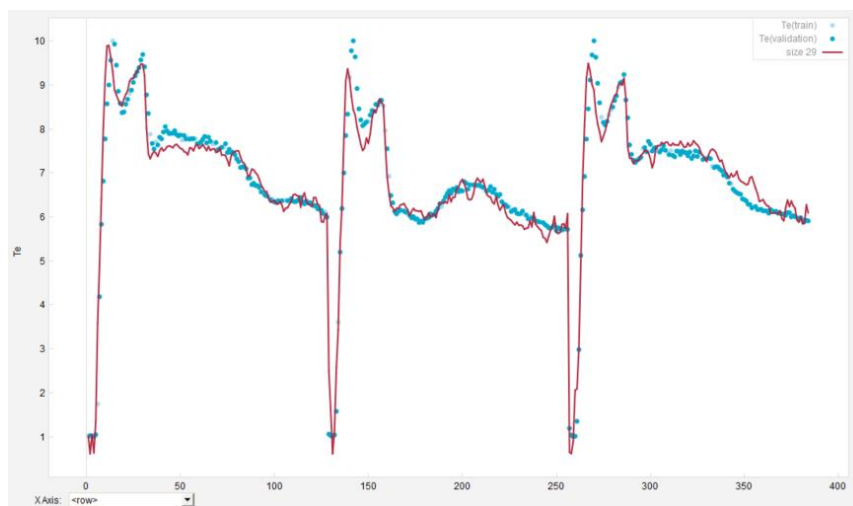


Ilustración 23. Curva del modelo vs Curva estimada

Una vez ya visualizado el modelo, se puede intuir que la complejidad ha sido elevada, no obstante, existe una gráfica que refleja en escalones de cuantificación el nivel de dificultad que ha causado al software una mayor o menor inversión de tiempo en generar el modelo.

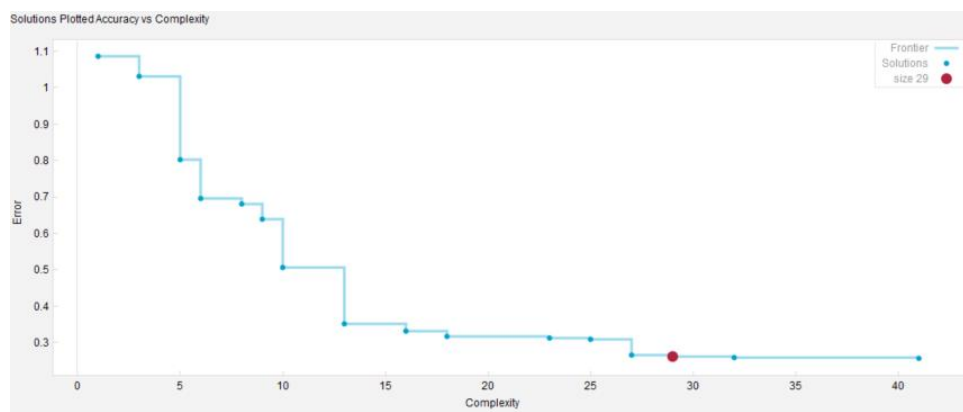


Ilustración 24. Curva error vs complejidad

Finalmente, en la pestaña “Reports”, se muestra la ecuación seleccionada para conocer el modelo matemático que describe el formato de dicha función. La cual podrá ser seleccionada y trasladarla a otro programa.

Model: $Te = a * Wp + b * Ha + c * Wp^3 + d * Wp * Ha * n^2 - e - f * Wp^2 - g * Ha * n^2$
Text: $Te = 5.00819032148707 * Wp + 0.722819959417676 * Ha + 0.0386153842111057 * Wp^3 + 0.000579997149239028 * Wp * Ha * n^2 - 4.36411039028033 - 0.801320512454427 * Wp^2 - 0.0100356284181321 * Ha * n^2$

Ilustración 25. Ecuación del modelo final

Una vez ya se obtiene dicha ecuación, se copia y traslada a la función de Matlab llamada `runEquation.m`, la cual se encarga de generar la curva real de la variable que hemos analizado y que realmente teníamos, con su curva diseñada mediante la técnica de regresión simbólica.

```

1  function runEquation();
2
3  load('discharge_raw.mat');
4
5  Te = discharge_raw(109).Te;
6  Ha = discharge_raw(109).Halp;
7  n = discharge_raw(109).n;
8  Ip = discharge_raw(109).Ip;
9  Wp = discharge_raw(109).Wp;
10
11
12  Te_pre= 5.00819032148707.*Wp + 0.722819959417676.*Ha + 0.0386153842111057.*Wp.^3 + 0.000579997149239028.*
13
14  mse = mean((Te_pre-Te).^2);
15  mae = mean(abs(Te_pre-Te));
16
17  figure;
18  hold on;
19  xlabel(sprintf('MSE %f, MAE %f',mse,mae));
20  plot(Te,'b');
21  plot(Te_pre,'r');
22  legend('Te','Te_pre');
```

Ilustración 26. Función runEquation.m

En la ilustración 25, se puede apreciar como en la línea 12 se define la variable `Te_pre`, la cual define Temperatura predicha, y en la línea 20 y 21 se ejecutan los plot de la curva `Te` y `Te_pre` con el fin de mostrar ambas curvas en una misma gráfica.

A su vez, con la línea 19 se refleja el Minimum Square Error y el Minimum Absolut Error, tomando principalmente el MSE como factor predominante para la comparativa y valoración de si el modelo definido es correcto y se adapta a la curva deseada.

Desarrollo Leyes de Potencia

La técnica analítica de leyes de potencia tiene como base principal y única el empleo del software de Matlab, el cual se basa en un principio en la obtención de las descargas similares al resto de variables analizadas en descargas previas, tal y como se explica el desarrollo en el apartado respectivo previo, donde se hace uso de las funciones “discharge raw.m” y “find previous discharges more similar to.m”.

Posteriormente, una vez se obtienen las descargas más similares de cada una de las otras cuatro variables de las cuales se conoce todos los valores que toma a lo largo del tiempo analizado, y por tanto la curva que describe, se podrá formar los modelos matemáticos incorporando exclusivamente las variables respectivas en cada caso.

La primera función por desarrollar es la de Ordinary Least Square, la cual se basa en la regresión lineal a partir del análisis de ciertas variables que permiten la cuantificación de un valor aplicándole el principio de los mínimos cuadrados. El código programado para el análisis de la variable temperatura con un tipo de calentamiento 1, resultaría del siguiente modo.

```
+4 run_PowerLaw_OLS.m x run_PowerLaw_RR.m x run_PowerLaw_TLS.m x run_PowerLaw_RTLS.m x
1 function coefs = run_PowerLaw_OLS(idTest, idTrain, idPredict);
2
3 load('discharge_raw.mat');
4
5 Te_test = discharge_raw(idTest).Te;
6 n_test = discharge_raw(idTest).n;
7 Wp_test = discharge_raw(idTest).Wp;
8 Ip_test = discharge_raw(idTest).Ip;
9 Ha_test = discharge_raw(idTest).Halpaha;
10
11 Te_train = [];
12 n_train = [];
13 Wp_train = [];
14 Ip_train = [];
15 Ha_train = [];
16
17 for i=1:length(idTrain);
18     Te_train = [Te_train; discharge_raw(idTrain(i)).Te];
19     n_train = [n_train; discharge_raw(idTrain(i)).n];
20     Wp_train = [Wp_train; discharge_raw(idTrain(i)).Wp];
21     Ip_train = [Ip_train; discharge_raw(idTrain(i)).Ip];
22     Ha_train = [Ha_train; discharge_raw(idTrain(i)).Halpaha];
23 end;
24 train = [Te_train n_train Wp_train Ip_train Ha_train];
25 test = [Te_test n_test Wp_test Ip_test Ha_test];
26
27 x_train = train;
28 x_test = test;
```

```

29
30 - x_train(:,idPredict) = [];
31 - x_test(:,idPredict) = [];
32 - y_train = train(:,idPredict);
33 - y_test = test(:,idPredict);
34
35 - yLog = log10(y_train);
36 - xLog = log10(x_train);
37
38 - b = ridge(yLog,xLog,0,0);
39
40 - y_predict = (10^b(1)).*(x_test(:,1).^b(2)).*(x_test(:,2).^b(3)).*(x_test(:,
41 - mse = mean((y_predict-y_test).^2);
42 - mae = mean(abs(y_predict-y_test));
43
44 - figure;
45 - hold on;
46
47 - if idPredict==1
48 -     plot(y_test,'Color',[0, 0.4470, 0.7410]);
49 -     plot(y_predict,'k');
50 -     legend('Te','Te pre');
51 - elseif idPredict==2
52 -     plot(y_test,'Color',[0.8500, 0.3250, 0.0980]);
53 -     plot(y_predict,'k');
54 -     legend('n','n pre');
55 - elseif idPredict==3
56 -     plot(y_test,'Color',[0.9290, 0.6940, 0.1250]);
57 -     plot(y_predict,'k');
58 -     legend('Wp','Wp pre');
59 - elseif idPredict==4
60 -     plot(y_test,'Color',[0.4940, 0.1840, 0.5560]);
61 -     plot(y_predict,'k');
62 -     legend('Ip','Ip pre');
63 - elseif idPredict==5
64 -     plot(y_test,'Color',[0.4660, 0.6740, 0.1880]);
65 -     plot(y_predict,'k');
66 -     legend('Ha','Ha pre');
67 - end;
68 - title('Ordinary Least Square');
69 - xlabel(sprintf('MSE %f, MAE %f',mse,mae));
70
71 - coefs = b;
72 - disp('_____');

```

Ilustración 27.Función OLS

A continuación, se establece el código referente a la función de Ridge Regression, la cual como se explico en su respectivo apartado, tiene como principal objetivo el análisis de la colinealidad que pueda existir entre las variables que diseñan el modelo matemático con el cual se pretende diseñar una función que permita demostrar la forma de la curva de la gráfica que tomara la variable que deseamos obtener, en función de las otras cuatro variables restantes, de las cuales se conocen todos sus valores. Dicho código tiene como base la programación procedente del libro escrito por Hoerl y Kennard [Hoerl et al,1970], aunque también cabe destacar que en Matlab existe una función que hace las veces de Ridge Regression, pero en este caso no se empleará.

```

+4 run_PowerLaw_OLS.m x run_PowerLaw_RR.m x run_PowerLaw_TLS.m x run_PowerLaw_RTLS.m x
1 function coefs = run_PowerLaw_RR(idTest, idTrain, idPredict);
2
3 load('discharge_raw.mat');
4
5 Te_test = discharge_raw(idTest).Te;
6 n_test = discharge_raw(idTest).n;
7 Wp_test = discharge_raw(idTest).Wp;
8 Ip_test = discharge_raw(idTest).Ip;
9 Ha_test = discharge_raw(idTest).Halp;
10
11 Te_train = [];
12 n_train = [];
13 Wp_train = [];
14 Ip_train = [];
15 Ha_train = [];
16
17 for i=1:length(idTrain);
18     Te_train = [Te_train; discharge_raw(idTrain(i)).Te];
19     n_train = [n_train; discharge_raw(idTrain(i)).n];
20     Wp_train = [Wp_train; discharge_raw(idTrain(i)).Wp];
21     Ip_train = [Ip_train; discharge_raw(idTrain(i)).Ip];
22     Ha_train = [Ha_train; discharge_raw(idTrain(i)).Halp];
23 end;
24 train = [Te_train n_train Wp_train Ip_train Ha_train];
25 test = [Te_test n_test Wp_test Ip_test Ha_test];
26
27 x_train = train;
28 x_test = test;
29
30 x_train(:,idPredict) = [];
31 x_test(:,idPredict) = [];
32 y_train = train(:,idPredict);
33 y_test = test(:,idPredict);
34
35 yLog = log10(y_train);
36 xLog = log10(x_train);
37
38 b=[];
39 y_predict = [];
40 for k=1:1000
41     b(:,k) = ridge(yLog,xLog,k,0);
42     y_predict_train(:,k) = (10^b(1,k)).*(x_train(:,1).^b(2,k)).*(x_train(:,2).^b(3,k));
43     mse_train(k) = mean((y_predict_train(:,k)-y_train).^2);
44 end;
45 [valor, indi] = min(mse_train);
46
47 y_predict = (10^b(1,indi)).*(x_test(:,1).^b(2,indi)).*(x_test(:,2).^b(3,indi));
48 mse = mean((y_predict-y_test).^2);
49 mae = mean(abs(y_predict-y_test));
50
51
52 figure;
53 hold on;
54

```



```

55 - if idPredict==1
56 -     plot(y_test,'Color',[0, 0.4470, 0.7410]);
57 -     plot(y_predict,'k');
58 -     legend('Te','Te pre');
59 - elseif idPredict==2
60 -     plot(y_test,'Color',[0.8500, 0.3250, 0.0980]);
61 -     plot(y_predict,'k');
62 -     legend('n','n pre');
63 - elseif idPredict==3
64 -     plot(y_test,'Color',[0.9290, 0.6940, 0.1250]);
65 -     plot(y_predict,'k');
66 -     legend('Wp','Wp pre');
67 - elseif idPredict==4
68 -     plot(y_test,'Color',[0.4940, 0.1840, 0.5560]);
69 -     plot(y_predict,'k');
70 -     legend('Ip','Ip pre');
71 - elseif idPredict==5
72 -     plot(y_test,'Color',[0.4660, 0.6740, 0.1880]);
73 -     plot(y_predict,'k');
74 -     legend('Ha','Ha pre');
75 - end;
76 - title('Ridge Regression');
77 - xlabel(sprintf('MSE %f, MAE %f, Ridge-Coeff %d',mse,mae, indi));
78 -
79 - coefs = b(:,indi);
80 - disp('_____');

```

Posteriormente, se establece el código referente a la técnica de Total Least Square en base a lo que se define en el libro escrito por Golub [Golub et al,1980], la cual se fundamenta teóricamente por ser bastante similar a la de Ordinary Least Square, es decir, su base principal de análisis tiene como origen el principio de los mínimos cuadrados.

```

1 - function coefs = run_PowerLaw_TLS(idTest, idTrain, idPredict);
2 -
3 - load('discharge_raw.mat');
4 -
5 - Te_test = discharge_raw(idTest).Te;
6 - n_test = discharge_raw(idTest).n;
7 - Wp_test = discharge_raw(idTest).Wp;
8 - Ip_test = discharge_raw(idTest).Ip;
9 - Ha_test = discharge_raw(idTest).Halp;
10 -
11 - Te_train = [];
12 - n_train = [];
13 - Wp_train = [];
14 - Ip_train = [];
15 - Ha_train = [];
16 -
17 - for i=1:length(idTrain)
18 -     n_train = [n_train; discharge_raw(idTrain(i)).n];
19 -     Wp_train = [Wp_train; discharge_raw(idTrain(i)).Wp];
20 -     Ip_train = [Ip_train; discharge_raw(idTrain(i)).Ip];
21 -     Ha_train = [Ha_train; discharge_raw(idTrain(i)).Halp];
22 -
23 - end;
24 - train = [Te_train n_train Wp_train Ip_train Ha_train];
25 - test = [Te_test n_test Wp_test Ip_test Ha_test];

```

Ilustración 28.Función Ridge Regression

```

25 - test = [Te_test n_test Wp_test Ip_test Ha_test];
26
27 - x_train = train;
28 - x_test = test;
29
30 - x_train(:,idPredict) = [];
31 - x_test(:,idPredict) = [];
32 - y_train = train(:,idPredict);
33 - y_test = test(:,idPredict);
34
35 - yLog = log10(y_train);
36 - xLog = log10(x_train);
37
38 - l1 = size(xLog,1);
39 - e1 = ones(l1,1);
40 - Xmedia = mean(xLog);
41 - Ymedia = mean(yLog);
42 - M = [Xmedia Ymedia];
43 - [U,S,V]=svd( [(xLog - e1*Xmedia), yLog - Ymedia] );
44 - N = (-1/V(end,end)) * V(:,end); % scaling
45 - b(1,1) = -M*N;
46 - b(2,1) = N(1);
47 - b(3,1) = N(2);
48 - b(4,1) = N(3);
49 - b(5,1) = N(4);
50 - y_predict = (10^b(1)).*(x_test(:,1).^b(2)).*(x_test(:,2).^b(3)).*(x_test(:,
51
52 - mse = mean((y_predict-y_test).^2);
53 - mae = mean(abs(y_predict-y_test));
54
55 - figure;
56 - hold on;
57
58 - if idPredict==1
59 -     plot(y_test,'Color',[0, 0.4470, 0.7410]);
60 -     plot(y_predict,'k');
61 -     legend('Te','Te pre');
62 - elseif idPredict==2
63 -     plot(y_test,'Color',[0.8500, 0.3250, 0.0980]);
64 -     plot(y_predict,'k');
65 -     legend('n','n pre');
66 - elseif idPredict==3
67 -     plot(y_test,'Color',[0.9290, 0.6940, 0.1250]);
68 -     plot(y_predict,'k');
69 -     legend('Wp','Wp pre');
70 - elseif idPredict==4
71 -     plot(y_test,'Color',[0.4940, 0.1840, 0.5560]);
72 -     plot(y_predict,'k');
73 -     legend('Ip','Ip pre');
74 - elseif idPredict==5
75 -     plot(y_test,'Color',[0.4660, 0.6740, 0.1880]);
76 -     plot(y_predict,'k');
77 -     legend('Ha','Ha pre');
78 - end;
79 - title('Total Least Square');
80 - xlabel(sprintf('MSE %f, MAE %f',mse,mae));
81

```

Ilustración 29.Funcion TLS

A la hora de programar el código de la técnica Regularized Total Least Square, lo que se pretende conseguir es conseguir una reducción del ruido en las gráficas que se obtengan de tal modo que su principal función será obtener resultados sin grandes oscilaciones en breves instantes de tiempo. Y esto se consigue mediante la programación partiendo de la base de lo que escribe Golub en su libro [Golub et al,1980b]

```
+5
run_PowerLaw_RR.m  run_PowerLaw_TLS.m  run_PowerLaw_RTLS.m  run_PowerLaw_RobustR.m
1  function coefs = run_PowerLaw_RTLS(idTest, idTrain, idPredict);
2
3  load('discharge_raw.mat');
4
5  Te_test = discharge_raw(idTest).Te;
6  n_test = discharge_raw(idTest).n;
7  Wp_test = discharge_raw(idTest).Wp;
8  Ip_test = discharge_raw(idTest).Ip;
9  Ha_test = discharge_raw(idTest).Halp;
10
11  Te_train = [];
12  n_train = [];
13  Wp_train = [];
14  Ip_train = [];
15  Ha_train = [];
16
17  for i=1:length(idTrain);
18      Te_train = [Te_train; discharge_raw(idTrain(i)).Te];
19      n_train = [n_train; discharge_raw(idTrain(i)).n];
20      Wp_train = [Wp_train; discharge_raw(idTrain(i)).Wp];
21      Ip_train = [Ip_train; discharge_raw(idTrain(i)).Ip];
22      Ha_train = [Ha_train; discharge_raw(idTrain(i)).Halp];
23  end;
24  train = [Te_train n_train Wp_train Ip_train Ha_train];
25  test = [Te_test n_test Wp_test Ip_test Ha_test];
26
27  x_train = train;
28  x_test = test;
29
30  x_train(:,idPredict) = [];
31  x_test(:,idPredict) = [];
32  y_train = train(:,idPredict);
33  y_test = test(:,idPredict);
34
35  yLog = log10(y_train);
36  xLog = log10(x_train);
37
38
39  l1 = size(xLog,1);
40  e1 = ones(l1,1);
41  Xmedia = mean(xLog);
42  Ymedia = mean(yLog);
43  M = [Xmedia Ymedia];
44  minSVD = min(svd( [(xLog - e1*Xmedia), yLog - Ymedia ] ));
45  lambdaSVD = -(minSVD^2);
46  Ly= std(yLog);
47  Lx= std(xLog);
```

```

48 % xLogNorm = (xLog - e1*Xmedia);
49 xLogNorm = zscore(xLog);
50 % yLogNorm = yLog - Ymedia;
51 yLogNorm = zscore(yLog);
52 n = size(xLog,2);
53 I = eye(n);
54 % tabla = [-0.4727:0.001:0];
55 w3 = inv(xLogNorm'*xLogNorm+lambdaSVD*I)*xLogNorm'*yLogNorm;
56 w3(1) = (Ly/Lx(1))*w3(1);
57 w3(2) = (Ly/Lx(2))*w3(2);
58 w3(3) = (Ly/Lx(3))*w3(3);
59 w3(4) = (Ly/Lx(4))*w3(4);
60 w0 = Ymedia - Xmedia*w3;
61 w3 = [w0;w3];
62 b = w3;
63 y_predict = (10^b(1)).*(x_test(:,1).^b(2)).*(x_test(:,2).^b(3)).*(x_test(:,3).^b(4));
64
65 mse = mean((y_predict-y_test).^2);
66 mae = mean(abs(y_predict-y_test));
67
68
69 figure;
70 hold on;
71 if idPredict==1
72     plot(y_test,'Color',[0, 0.4470, 0.7410]);
73     plot(y_predict,'k');
74     legend('Te','Te pre');
75 elseif idPredict==2
76     plot(y_test,'Color',[0.8500, 0.3250, 0.0980]);
77     plot(y_predict,'k');
78     legend('n','n pre');
79 elseif idPredict==3
80     plot(y_test,'Color',[0.9290, 0.6940, 0.1250]);
81     plot(y_predict,'k');
82     legend('Wp','Wp pre');
83 elseif idPredict==4
84     plot(y_test,'Color',[0.4940, 0.1840, 0.5560]);
85     plot(y_predict,'k');
86     legend('Ip','Ip pre');
87 elseif idPredict==5
88     plot(y_test,'Color',[0.4660, 0.6740, 0.1880]);
89     plot(y_predict,'k');
90     legend('Ha','Ha pre');
91 end;
92 title('Regularized Total Least Square');
93 xlabel(sprintf('MSE %f, MAE %f',mse,mae));
94
95 coefs = b;
96 disp('_____');

```

Ilustración 30.Función RTLS

Y, por último, a la hora de definir el código referente a la función de Robust Regression, obtenido a partir del análisis del libro de P. Holland [Holland,1977], lo que se pretende conseguir es el diseño de un modelo mediante el análisis de todos los valores existentes, incluso aquellos que difieren en gran medida de la realidad, donde existen puntos tomados a lo largo del período a estudiar, los cuales son existentes.

```

+5  run_PowerLaw_RR.m  run_PowerLaw_TLS.m  run_PowerLaw_RTLS.m  run_PowerLaw_RobustR.m
1  function coefs = run_PowerLaw_RobustR(idTest, idTrain, idPredict);
2
3  load('discharge_raw.mat');
4
5  Te_test = discharge_raw(idTest).Te;
6  n_test = discharge_raw(idTest).n;
7  Wp_test = discharge_raw(idTest).Wp;
8  Ip_test = discharge_raw(idTest).Ip;
9  Ha_test = discharge_raw(idTest).Halp;
10
11 Te_train = [];
12 n_train = [];
13 Wp_train = [];
14 Ip_train = [];
15 Ha_train = [];
16
17 for i=1:length(idTrain);
18     Te_train = [Te_train; discharge_raw(idTrain(i)).Te];
19     n_train = [n_train; discharge_raw(idTrain(i)).n];
20     Wp_train = [Wp_train; discharge_raw(idTrain(i)).Wp];
21     Ip_train = [Ip_train; discharge_raw(idTrain(i)).Ip];
22     Ha_train = [Ha_train; discharge_raw(idTrain(i)).Halp];
23 end;
24 train = [Te_train n_train Wp_train Ip_train Ha_train];
25 test = [Te_test n_test Wp_test Ip_test Ha_test];
26
27 x_train = train;
28 x_test = test;
29
30 x_train(:,idPredict) = [];
31 x_test(:,idPredict) = [];
32 y_train = train(:,idPredict);
33 y_test = test(:,idPredict);
34
35 yLog = log10(y_train);
36 xLog = log10(x_train);
37
38
39 b(5,8)=0;
40 nombre(1).cadena = 'bisure';
41 b(:,1) = robustfit(xLog,yLog);
42 y_predict_train(:,1) = (10^b(1,1)).*(x_train(:,1).^b(2,1)).*(x_train(:,
43 mse(1) = mean((y_predict_train(:,1)-y_train).^2);
44 mae(1) = mean(abs(y_predict_train(:,1)-y_train));
45
46 nombre(2).cadena = 'andrews';
47 b(:,2) = robustfit(xLog,yLog,'andrews');
48 y_predict_train(:,2) = (10^b(1,2)).*(x_train(:,1).^b(2,2)).*(x_train(:,
49 mse(2) = mean((y_predict_train(:,2)-y_train).^2);
50 mae(2) = mean(abs(y_predict_train(:,2)-y_train));
51
52 nombre(2).cadena = 'cauchy';
53 b(:,3) = robustfit(xLog,yLog,'cauchy');
54 y_predict_train(:,3) = (10^b(1,3)).*(x_train(:,1).^b(2,3)).*(x_train(:,
55 mse(3) = mean((y_predict_train(:,3)-y_train).^2);

```

```

55 - mse(3) = mean((y_predict_train(:,3)-y_train).^2);
56 - mae(3) = mean(abs(y_predict_train(:,3)-y_train));
57
58 - nombre(4).cadena = 'fair';
59 - b(:,4) = robustfit(xLog,yLog,'fair');
60 - y_predict_train(:,4) = (10^b(1,4)).*(x_train(:,1).^b(2,4)).*(x_train(:,4)-y_train).^2);
61 - mse(4) = mean((y_predict_train(:,4)-y_train).^2);
62 - mae(4) = mean(abs(y_predict_train(:,4)-y_train));
63
64 - nombre(5).cadena = 'huber';
65 - b(:,5) = robustfit(xLog,yLog,'huber');
66 - y_predict_train(:,5) = (10^b(1,5)).*(x_train(:,1).^b(2,5)).*(x_train(:,5)-y_train).^2);
67 - mse(5) = mean((y_predict_train(:,5)-y_train).^2);
68 - mae(5) = mean(abs(y_predict_train(:,5)-y_train));
69
70 - nombre(6).cadena = 'logistic';
71 - b(:,6) = robustfit(xLog,yLog,'logistic');
72 - y_predict_train(:,6) = (10^b(1,6)).*(x_train(:,1).^b(2,6)).*(x_train(:,6)-y_train).^2);
73 - mse(6) = mean((y_predict_train(:,6)-y_train).^2);
74 - mae(6) = mean(abs(y_predict_train(:,6)-y_train));
75
76 - nombre(7).cadena = 'talwar';
77 - b(:,7) = robustfit(xLog,yLog,'talwar');
78 - y_predict_train(:,7) = (10^b(1,7)).*(x_train(:,1).^b(2,7)).*(x_train(:,7)-y_train).^2);
79 - mse(7) = mean((y_predict_train(:,7)-y_train).^2);
80 - mae(7) = mean(abs(y_predict_train(:,7)-y_train));
81
82 - nombre(8).cadena = 'welsch';
83 - b(:,8) = robustfit(xLog,yLog,'welsch');
84 - y_predict_train(:,8) = (10^b(1,8)).*(x_train(:,1).^b(2,8)).*(x_train(:,8)-y_train).^2);
85 - mse(8) = mean((y_predict_train(:,8)-y_train).^2);
86 - mae(8) = mean(abs(y_predict_train(:,8)-y_train));
87
88 - indiMin = find(mse==min(mse));
89
90 - y_predict = (10^b(1,indiMin(1))).*(x_test(:,1).^b(2,indiMin(1))).*(x_test(:,indiMin(1))-y_test).^2);
91 - mse = mean((y_predict-y_test).^2);
92 - mae = mean(abs(y_predict-y_test));
93
94 - figure;
95 - hold on;
96
97 - if idPredict==1
98 -     plot(y_test,'Color',[0, 0.4470, 0.7410]);
99 -     plot(y_predict,'k');
100 -     legend('Te','Te pre');
101 - elseif idPredict==2
102 -     plot(y_test,'Color',[0.8500, 0.3250, 0.0980]);
103 -     plot(y_predict,'k');
104 -     legend('n','n pre');
105 - elseif idPredict==3
106 -     plot(y_test,'Color',[0.9290, 0.6940, 0.1250]);
107 -     plot(y_predict,'k');
108 -     legend('Wp','Wp pre');

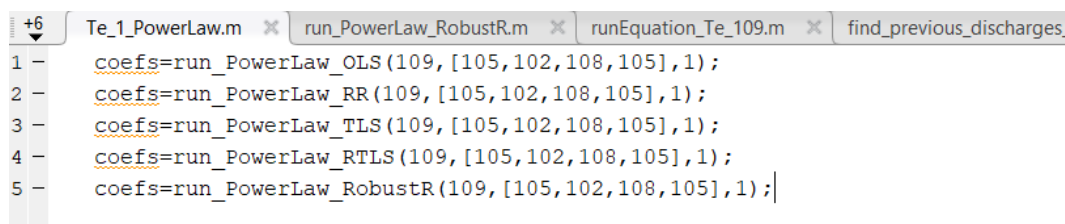
```

```

108 -     legend('Wp','Wp pre');
109 - elseif idPredict==4
110 -     plot(y_test,'Color',[0.4940, 0.1840, 0.5560]);
111 -     plot(y_predict,'k');
112 -     legend('Ip','Ip pre');
113 - elseif idPredict==5
114 -     plot(y_test,'Color',[0.4660, 0.6740, 0.1880]);
115 -     plot(y_predict,'k');
116 -     legend('Ha','Ha pre');
117 - end;
118 - title('Robust Regression');
119 - xlabel(sprintf('MSE %f,      MAE %f,      Weight-Function "%s"',mse,mae, nomb.
120 -
121 - coefs = b(:,indiMin(1));
122 - disp('_____');
```

Ilustración 31.Función Robust Regression

Finalmente, una vez se obtiene cada una de las funciones que programa el código referente a cada una de las técnicas desarrolladas, se generará un script que permitirá ejecutar la misma variable desde las distintas técnicas aportando una mejor capacidad a la hora de visualizar los resultados en la práctica.



```

+6 Te_1_PowerLaw.m x run_PowerLaw_RobustR.m x runEquation_Te_109.m x find_previous_discharges.
1 - coefs=run_PowerLaw_OLS(109,[105,102,108,105],1);
2 - coefs=run_PowerLaw_RR(109,[105,102,108,105],1);
3 - coefs=run_PowerLaw_TLS(109,[105,102,108,105],1);
4 - coefs=run_PowerLaw_RTLS(109,[105,102,108,105],1);
5 - coefs=run_PowerLaw_RobustR(109,[105,102,108,105],1);
```

Ilustración 32.Script final

En la imagen superior, se puede comprobar como a cada variable “coefs” se le asocia cada una de las técnicas empleadas. El primer término dentro del paréntesis es el número de descarga asociado a la lista que presenta en la base de datos que se ha utilizado. Una vez se inicia el corchete, se mencionan la descarga más similar a la descarga por analizar, en cada una de las variables que condicionan el modelo del sistema a formar. Y por último, antes de finalizar el paréntesis, el término que aparece hace referencia a la variable que se está analizando, es decir, 1 si es temperatura, 2 si es densidad, 3 si es Energía diamagnética, 4 si es corriente del plasma y 5 si es H-alpha.

A modo de conclusión, en el siguiente apartado una vez se ejecuten todos los scripts referentes a las leyes de potencia, se escogerá del total de las cinco técnicas por variable a analizar, la técnica que presente un menor error mínimo cuadrado o menor error absoluto, aunque siempre se cumple que la misma técnica presenta el error mínimo en ambos valores en comparación con el resto.

Resultados prácticos

En este apartado se pretende reflejar la diferencia a modo visual entre la gráfica obtenida mediante la técnica de regresión simbólica y la técnica de leyes de potencia, en la cual esta última mostrará únicamente el modelo que presente un menor error, tanto cuadrado como absoluto.

Por tanto, se realizará el análisis comparativo dividiendo el ensayo en función del tipo de calentamiento y de la variable a estudiar en cada caso.

Calentamiento tipo 1

Este tipo de calentamiento se ha analizado respecto de la descarga 33308, o bien la descarga número 108 de la base de datos con la que se realiza el estudio.

Variable T_e

Mediante el análisis de la variable temperatura en el calentamiento de tipo 1, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

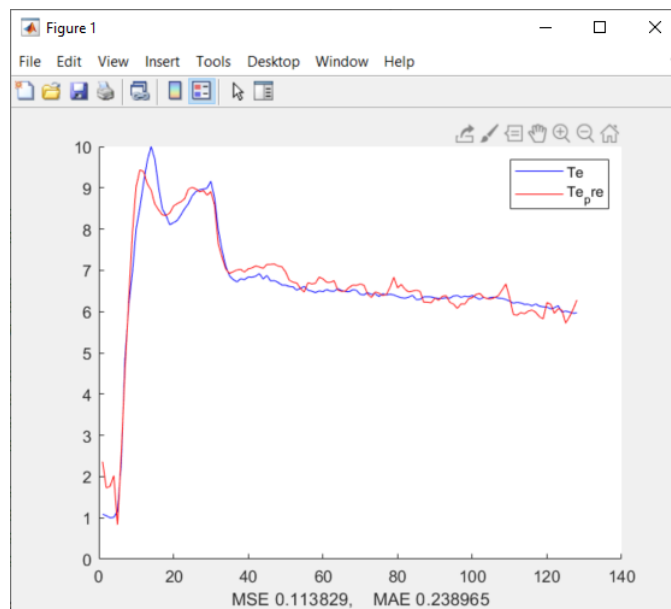


Ilustración 33. T_e Regresión simbólica Calentamiento 1

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regresión.

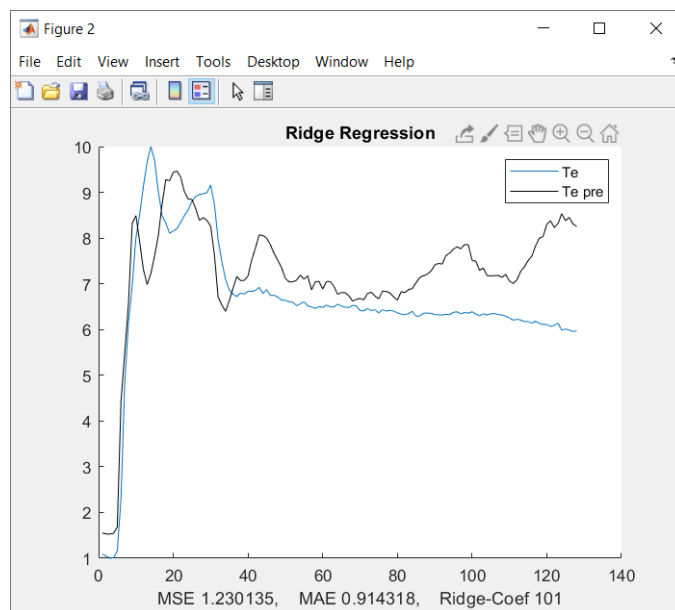


Ilustración 34. Te Leyes de potencia Calentamiento 1

Variable n

Mediante el análisis de la variable densidad en el calentamiento de tipo 1, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

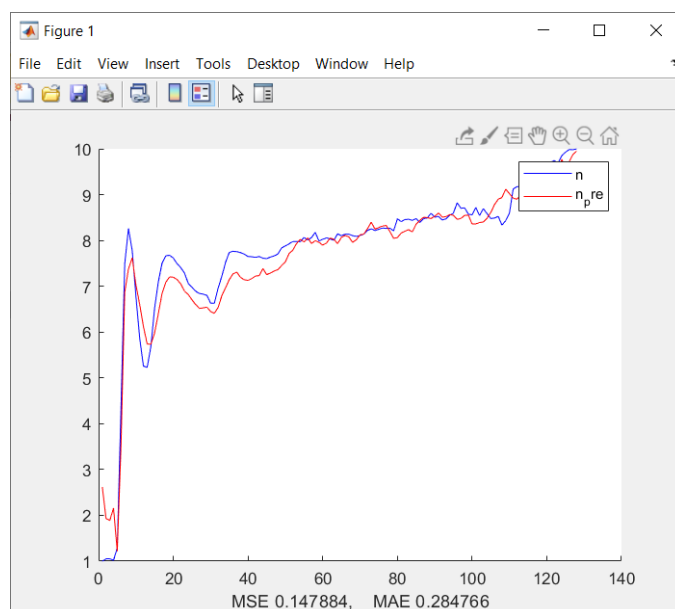


Ilustración 35. N Regresión simbólica Calentamiento 1

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Robust Regresión.

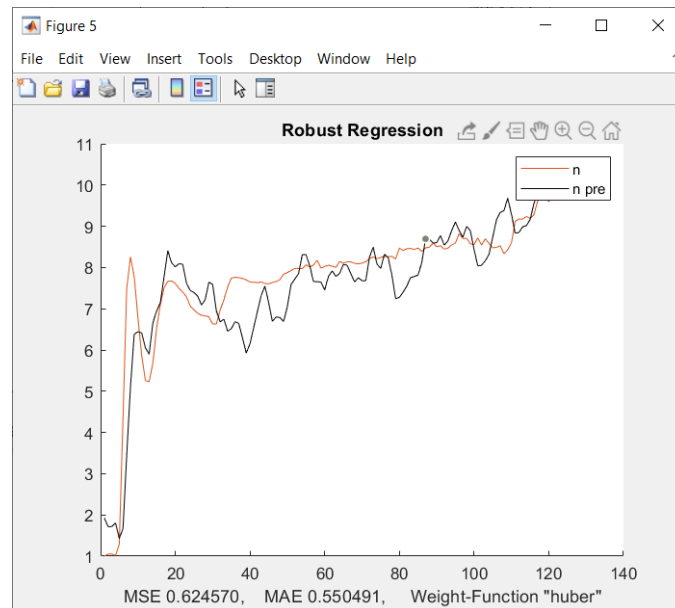


Ilustración 36. N Leyes de potencia Calentamiento 1

Variable W_p

Mediante el análisis de la variable Energía diamagnética en el calentamiento de tipo 1, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

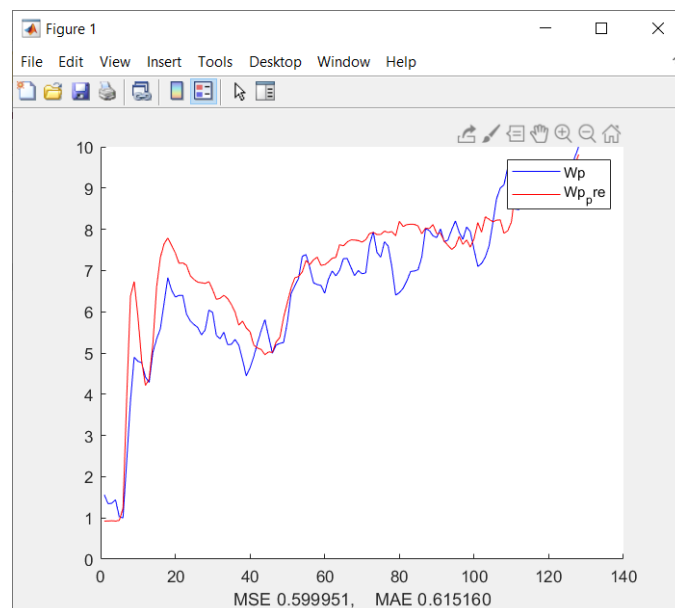


Ilustración 37. Wp Regresión simbólica Calentamiento 1

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Robust Regresión.

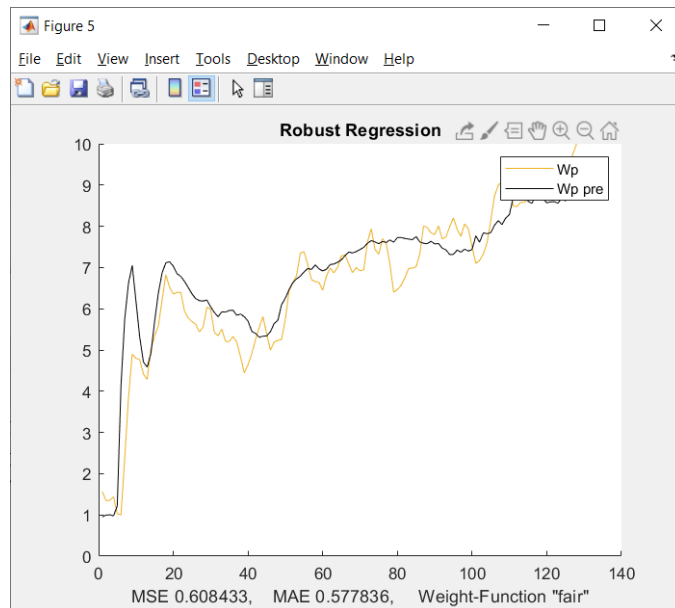


Ilustración 38. Wp Leyes de potencia Calentamiento 1

Variable I_p

Mediante el análisis de la variable corriente del plasma en el calentamiento de tipo 1, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

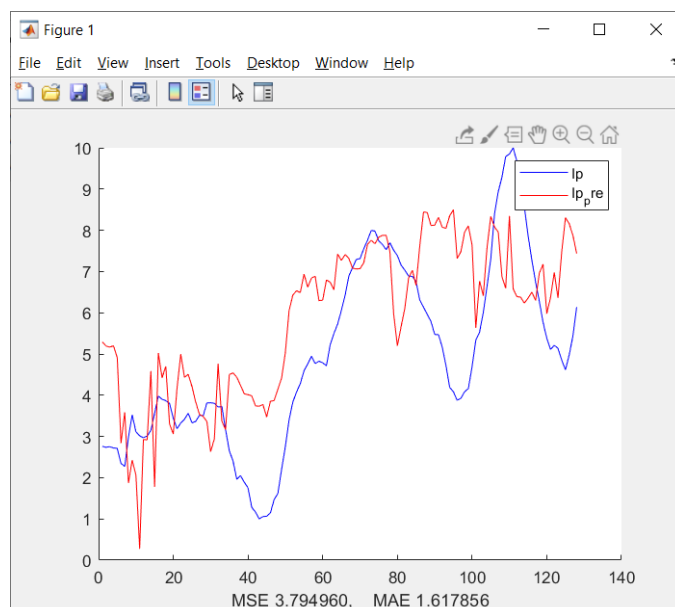


Ilustración 39. I_p Regresión simbólica Calentamiento 1

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

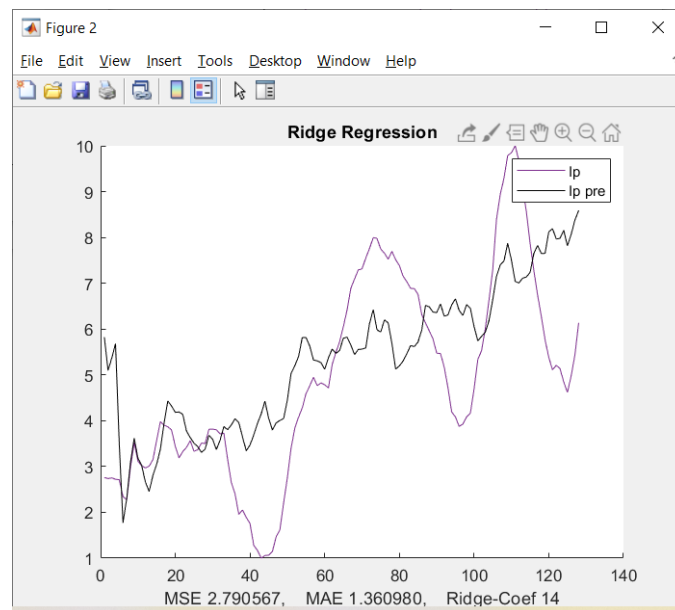


Ilustración 40. Ip Leyes de potencia Calentamiento 1

Variable Ha

Mediante el análisis de la variable H-alpha en el calentamiento de tipo 1, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

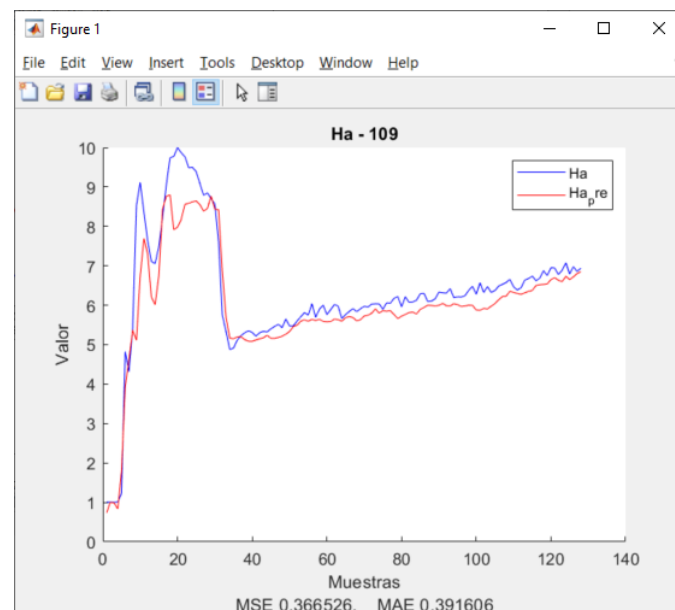


Ilustración 41. Ha Regresión simbólica Calentamiento 1

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Regularized Total Least Square.

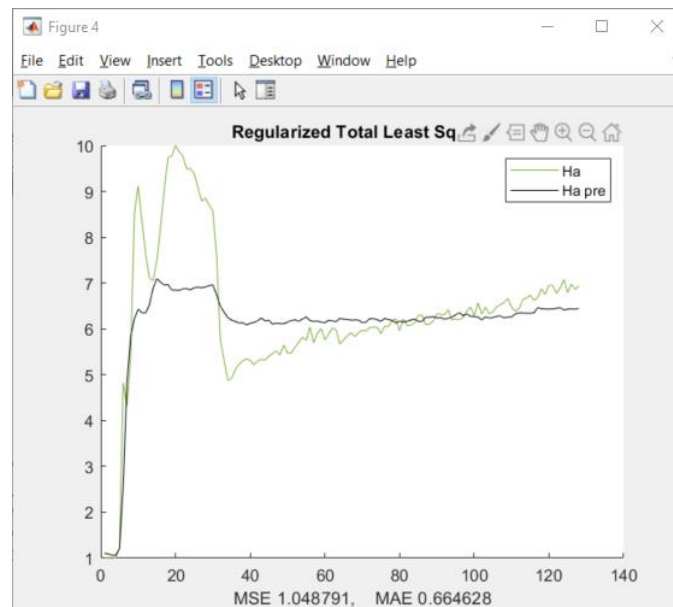


Ilustración 42. Ha Leyes de potencia Calentamiento 1

Calentamiento tipo 2

Este tipo de calentamiento se ha analizado respecto de la descarga 44485, o bien la descarga número 422 de la base de datos con la que se realiza el estudio.

Variable T_e

Mediante el análisis de la variable temperatura en el calentamiento de tipo 2, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

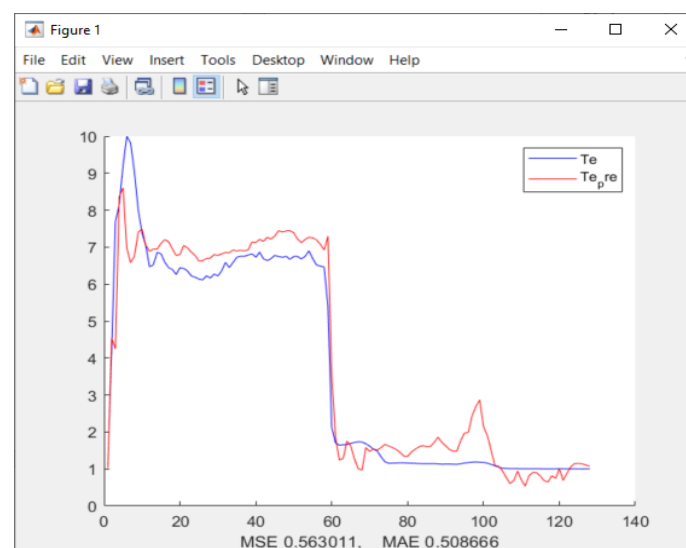


Ilustración 43. Te Regresión simbólica Calentamiento 2

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Total Least Square.

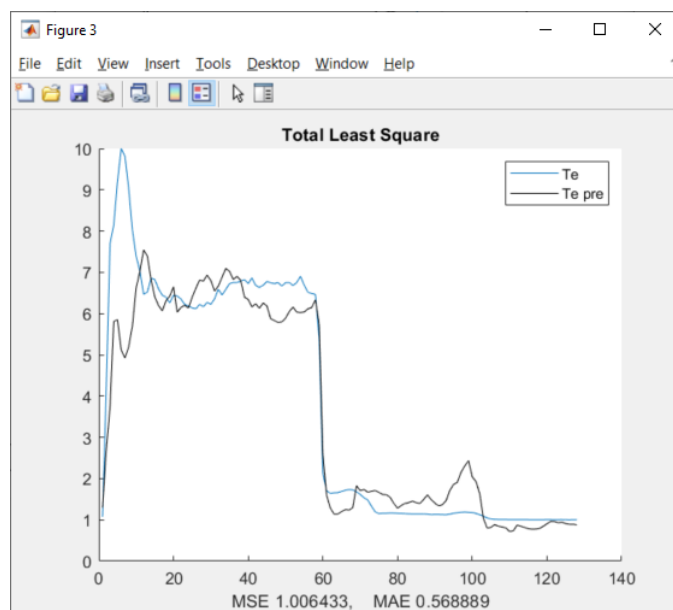


Ilustración 44. T_e Leyes de potencia Calentamiento 2

Variable n

Mediante el análisis de la variable densidad en el calentamiento de tipo 2, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

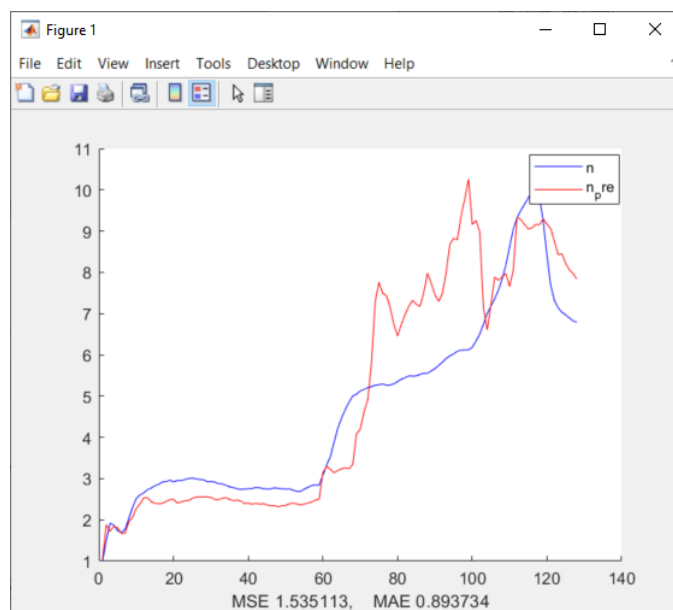


Ilustración 45. N Regresión simbólica Calentamiento 2

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Robust Regression.

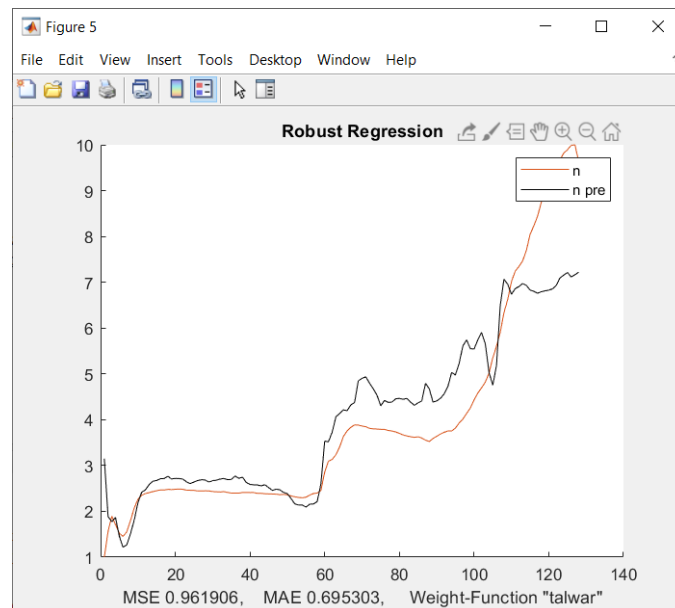


Ilustración 46. N Leyes de potencia Calentamiento 2

Variable W_p

Mediante el análisis de la variable W_p en el calentamiento de tipo 2, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

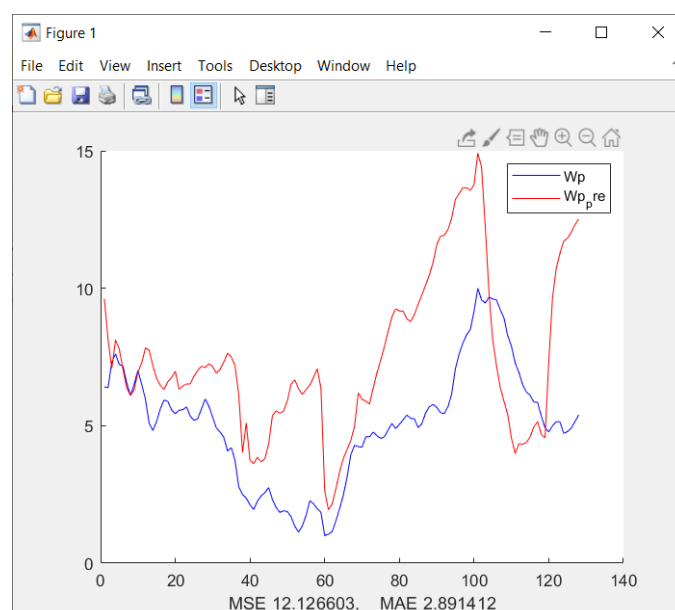


Ilustración 47. Wp Regresión simbólica Calentamiento 2

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Robust Regression.

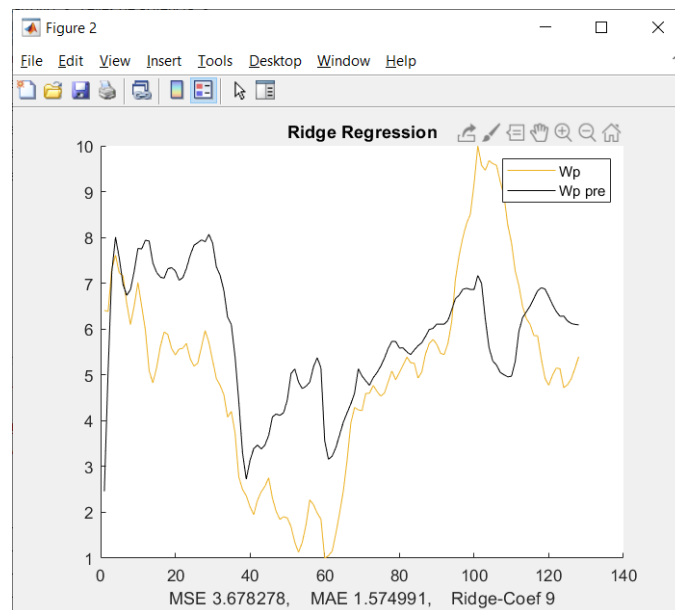


Ilustración 48. Wp Leyes de potencia Calentamiento 2

Variable I_p

Mediante el análisis de la variable I_p en el calentamiento de tipo 2, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

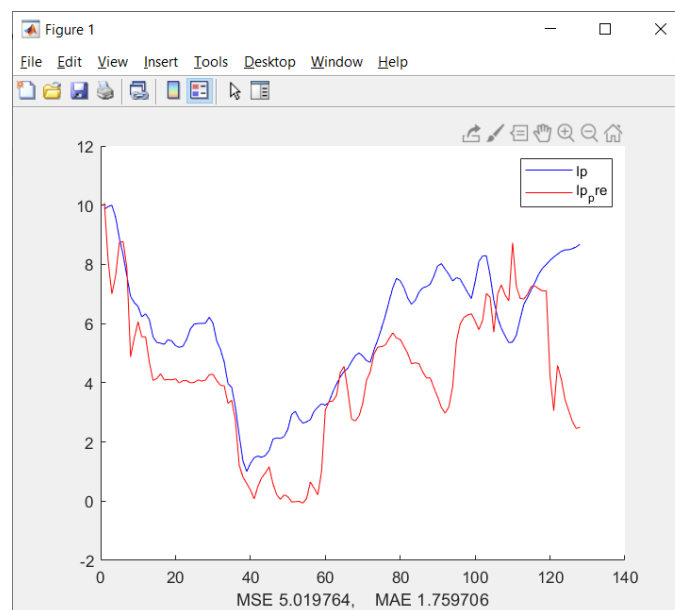


Ilustración 49. Ip Regresión simbólica Calentamiento 2

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

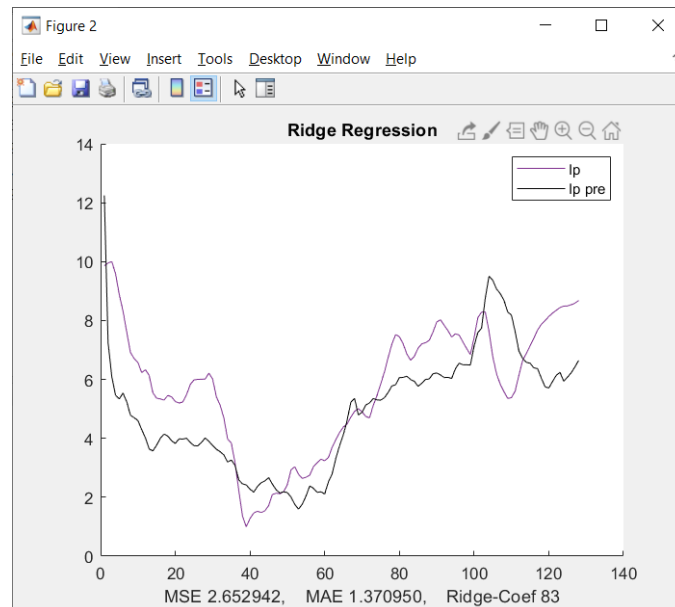


Ilustración 50. Ip Leyes de potencia Calentamiento 2

Variable Ha

Mediante el análisis de la variable H-alpha en el calentamiento de tipo 2, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

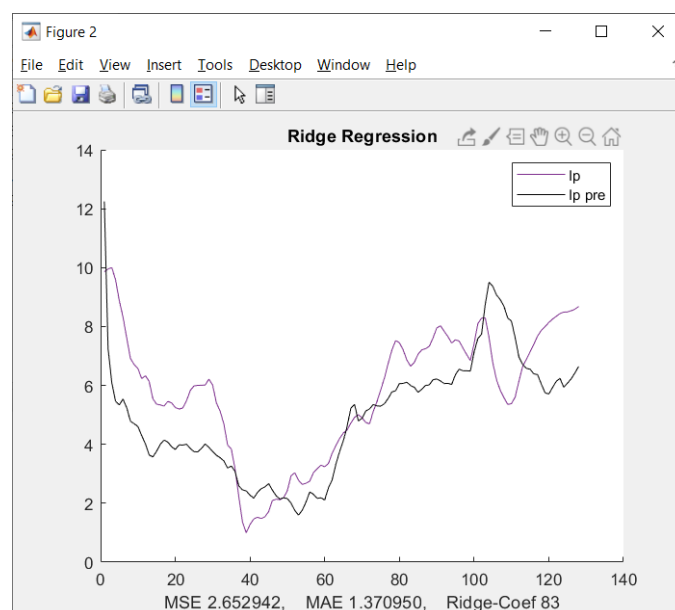


Ilustración 51. Ha Regresión simbólica Calentamiento 2

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Robust Regression.

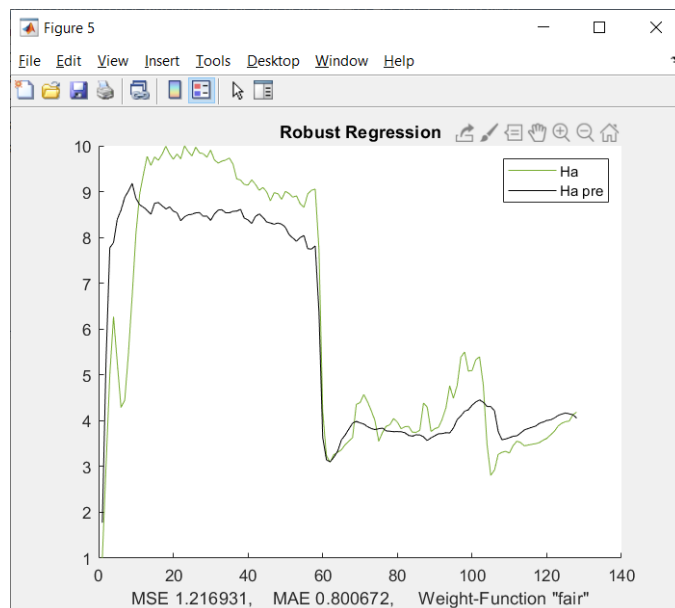


Ilustración 52. Ha Leyes de potencia Calentamiento 2

Calentamiento tipo 3

Este tipo de calentamiento se ha analizado respecto de la descarga 33032, o bien la descarga número 92 de la base de datos con la que se realiza el estudio.

Variable T_e

Mediante el análisis de la variable temperatura en el calentamiento de tipo 3, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

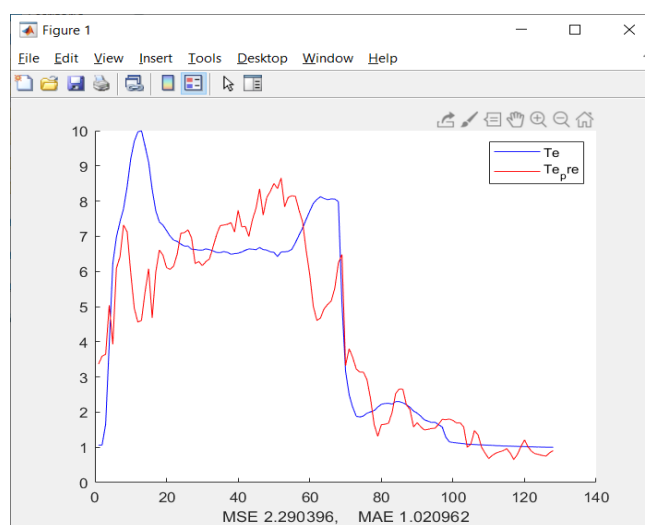


Ilustración 53. Te Regresión simbólica Calentamiento 3

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Robust Regression.

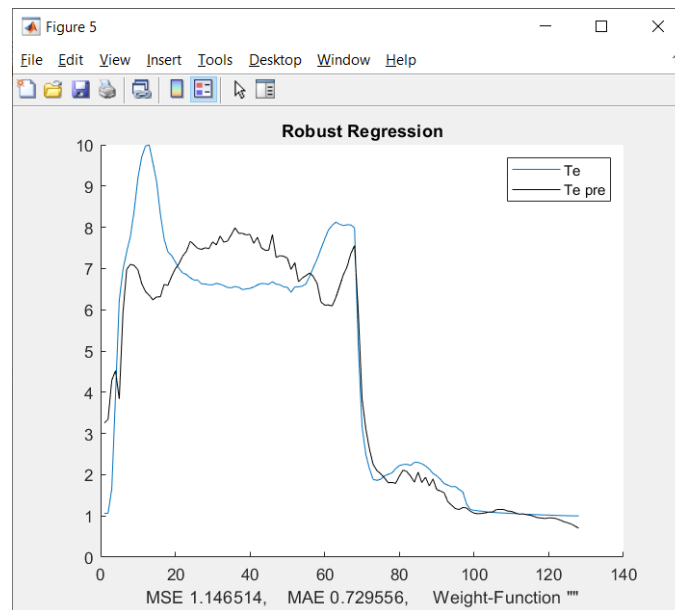


Ilustración 54. Te Leyes de potencia Calentamiento 3

Variable n

Mediante el análisis de la variable densidad en el calentamiento de tipo 3, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

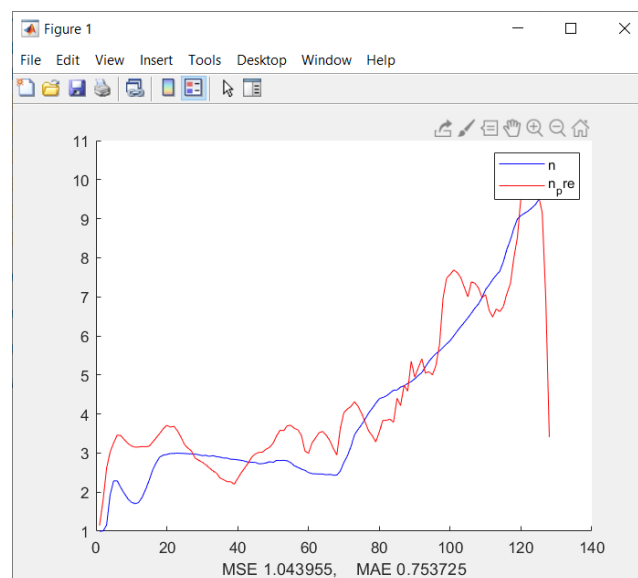


Ilustración 55. N Regresión simbólica Calentamiento 3

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Robust Regression.

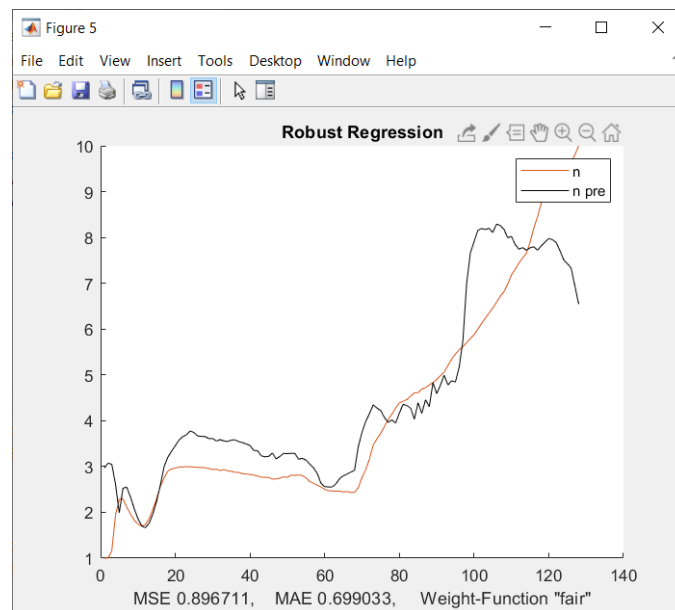


Ilustración 56. N Leyes de potencia Calentamiento 3

Variable Wp

Mediante el análisis de la variable Wp en el calentamiento de tipo 3, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

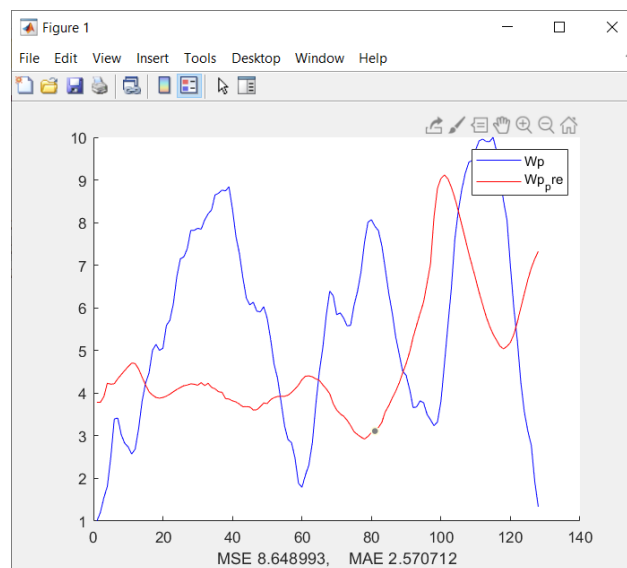


Ilustración 57. Wp Regresión simbólica Calentamiento 3

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Robust Regression.

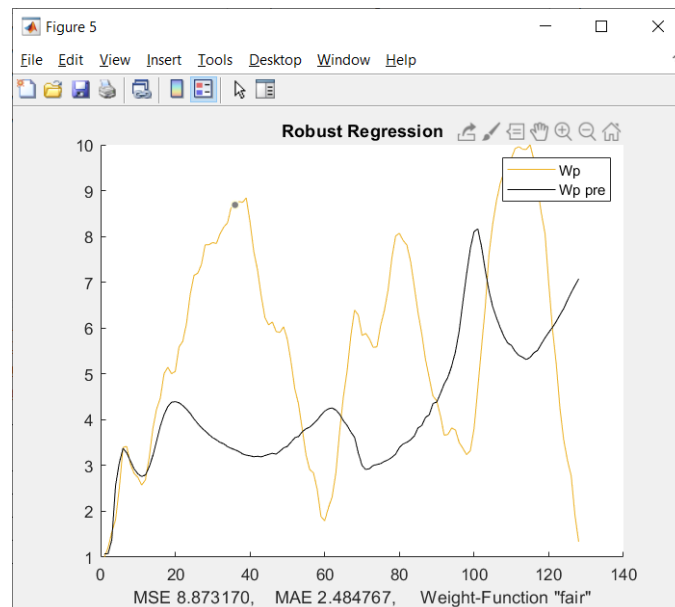


Ilustración 58. Wp Leyes de potencia Calentamiento 3

Variable I_p

Mediante el análisis de la variable I_p en el calentamiento de tipo 3, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

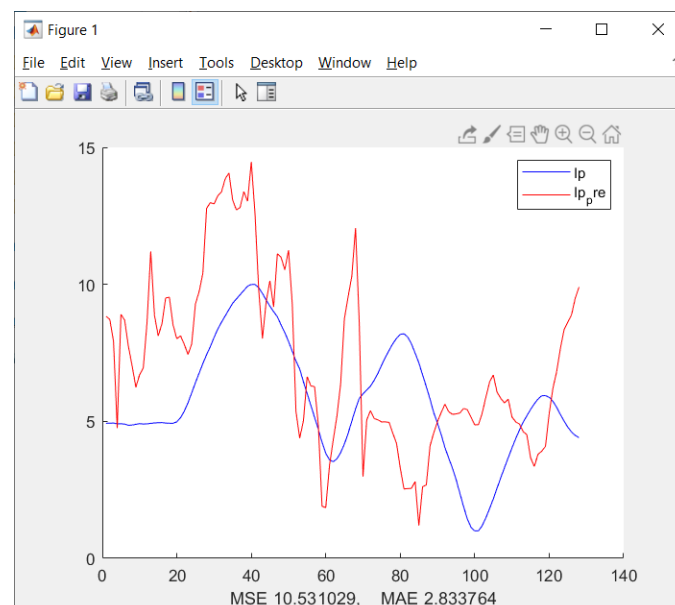


Ilustración 59. I_p Regresión simbólica Calentamiento 3

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

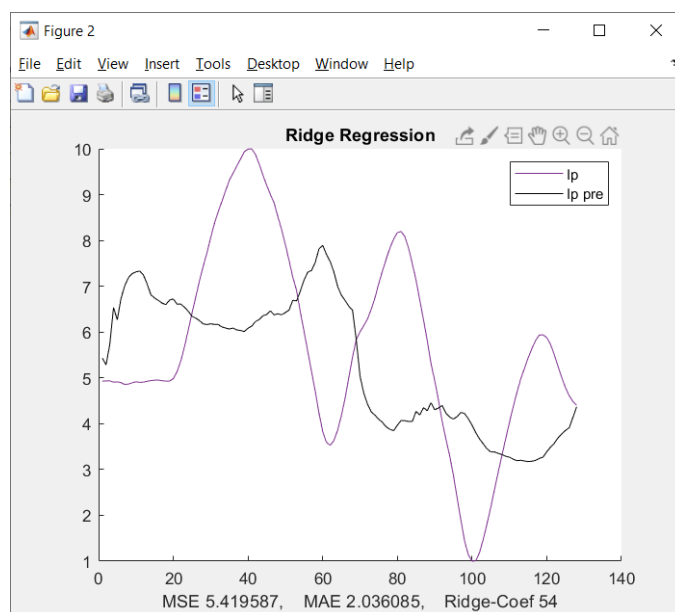


Ilustración 60. Ip Leyes de potencia Calentamiento 3

Variable H_a

Mediante el análisis de la variable H-alpha en el calentamiento de tipo 3, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

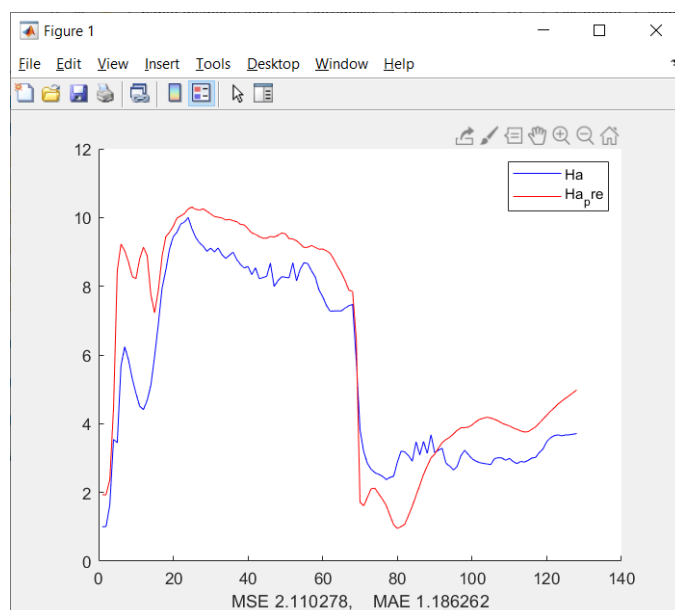


Ilustración 61. Ha Regresión simbólica Calentamiento 3

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Robust Regression.

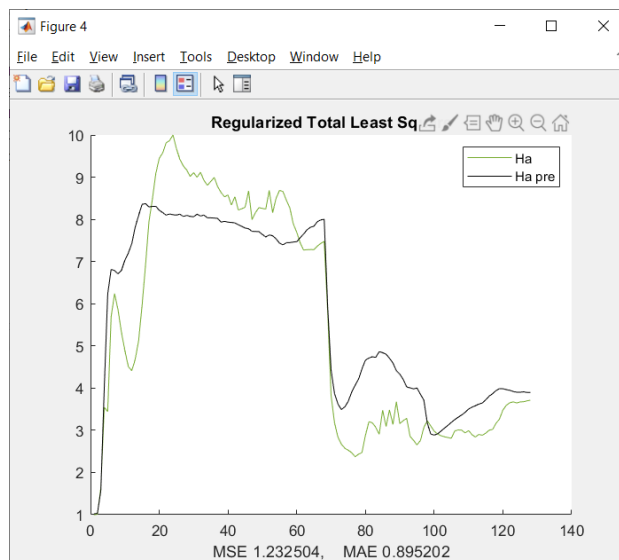


Ilustración 62. Ha Leyes de potencia Calentamiento 3

Calentamiento tipo 4

Este tipo de calentamiento se ha analizado respecto de la descarga 37541, o bien la descarga número 220 de la base de datos con la que se realiza el estudio.

Variable T_e

Mediante el análisis de la variable temperatura en el calentamiento de tipo 4, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Total Least Square.

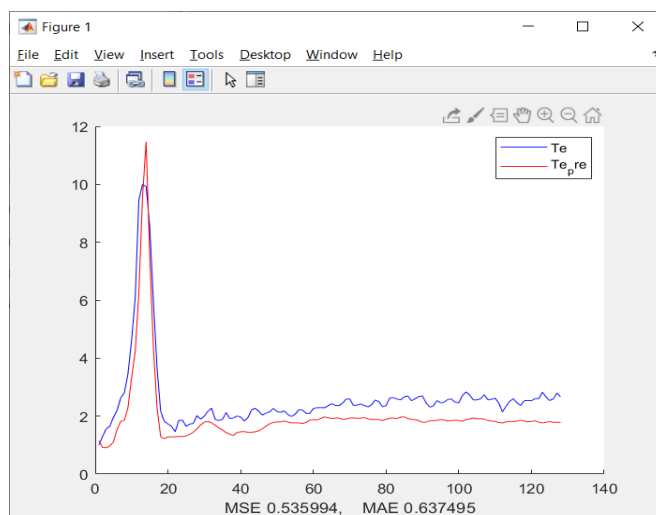


Ilustración 63. Te Regresión simbólica Calentamiento 4

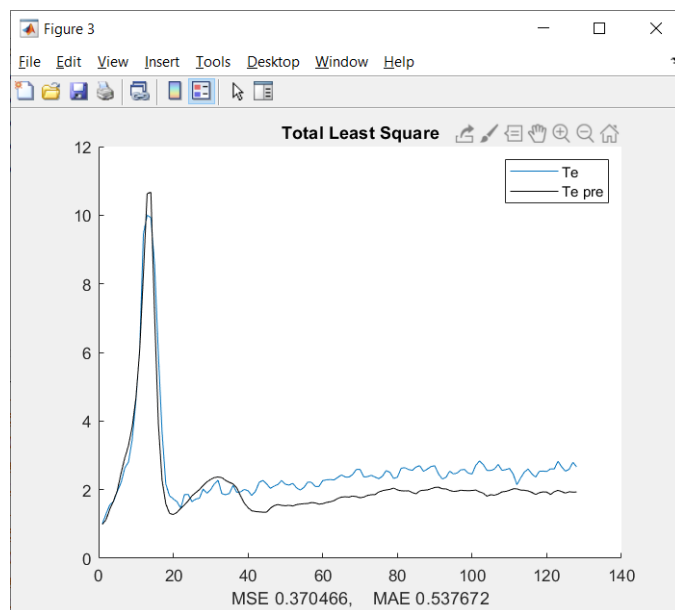


Ilustración 64. T_e Leyes de potencia Calentamiento 4

Variable n

Mediante el análisis de la variable densidad en el calentamiento de tipo 4, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

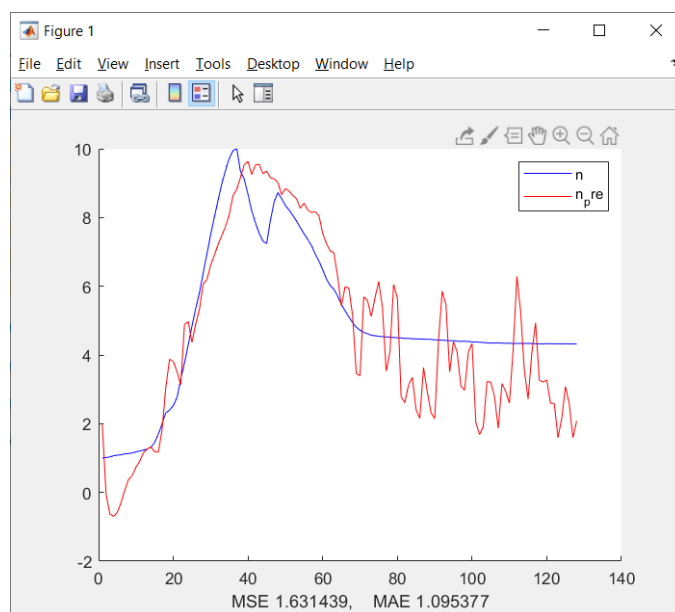


Ilustración 65. N Regresión simbólica Calentamiento 4

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

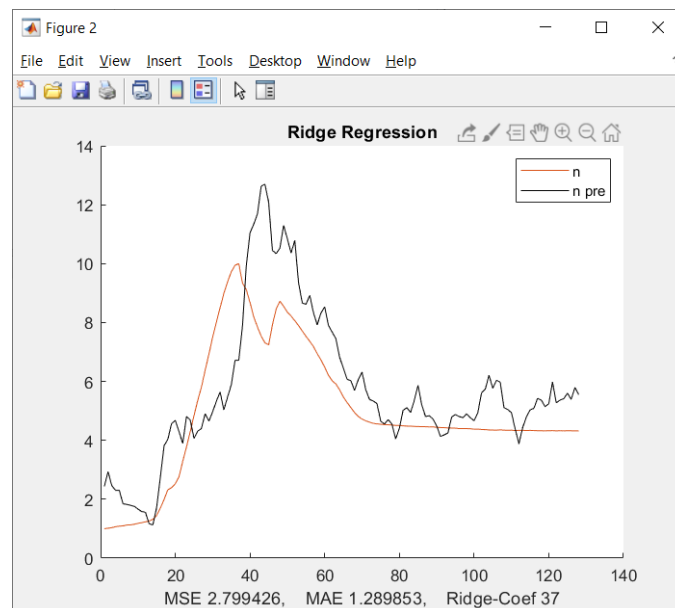


Ilustración 66. N Leyes de potencia Calentamiento 4

Variable W_p

Mediante el análisis de la variable W_p en el calentamiento de tipo 4, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

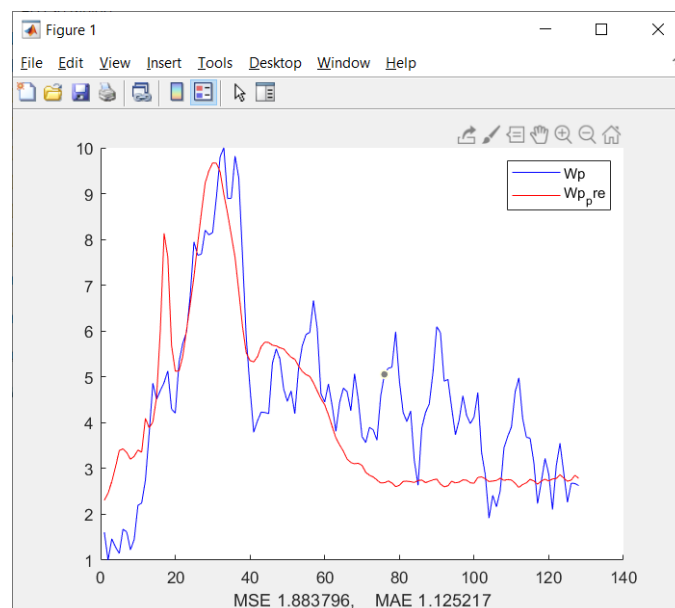


Ilustración 67. W_p Regresión simbólica Calentamiento 4

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

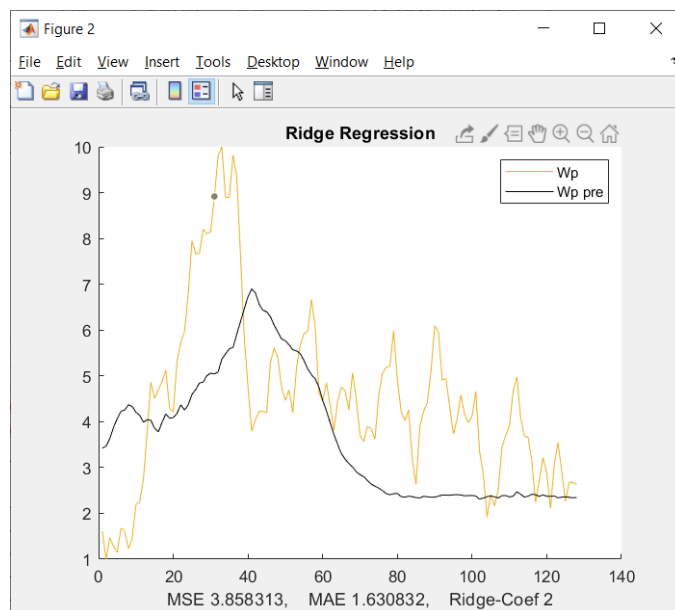


Ilustración 68. Wp Leyes de potencia Calentamiento 4

Variable I_p

Mediante el análisis de la variable I_p en el calentamiento de tipo 4, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

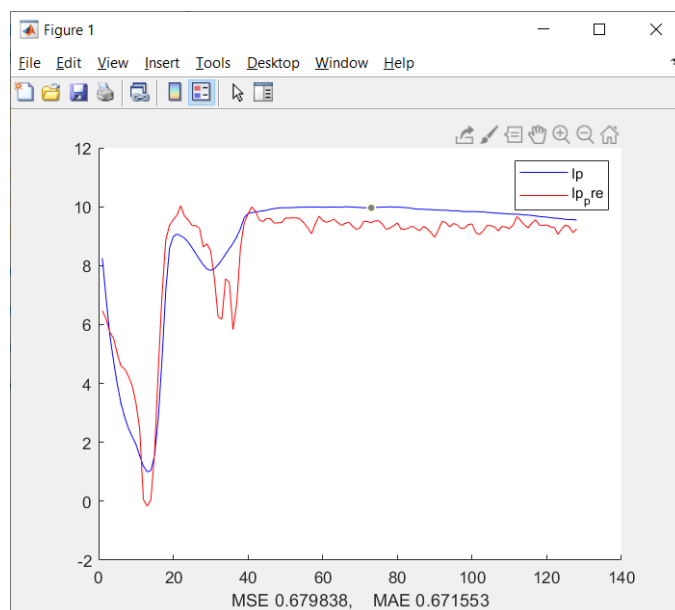


Ilustración 69. I_p Regresión simbólica Calentamiento 4

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Regularized Total Least Square.

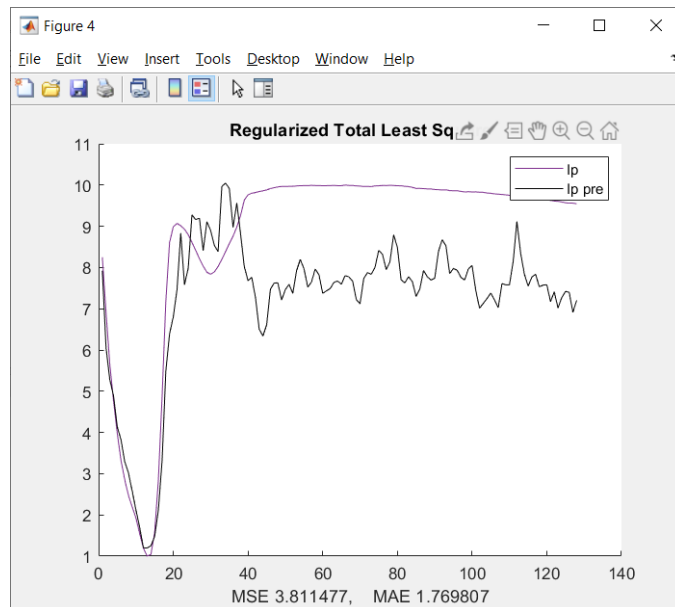


Ilustración 70. Ip Leyes de potencia Calentamiento 4

Variable H_α

Mediante el análisis de la variable H-alpha en el calentamiento de tipo 4, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

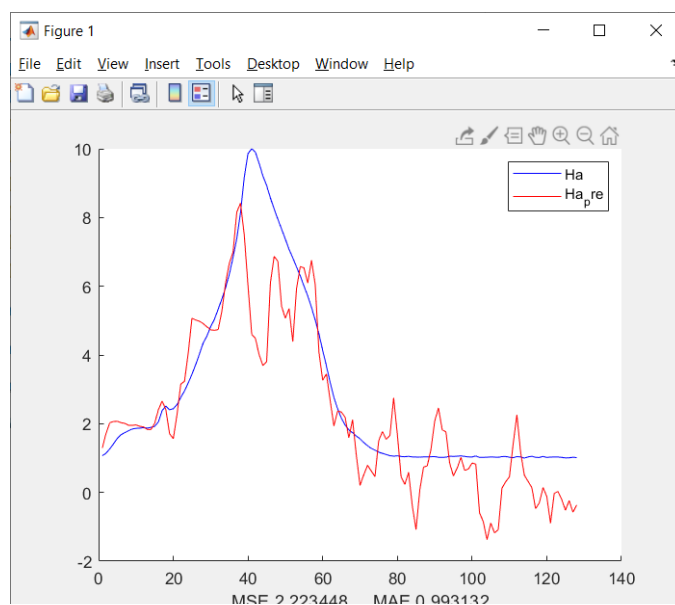


Ilustración 71. Ha Regresión simbólica Calentamiento 4

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

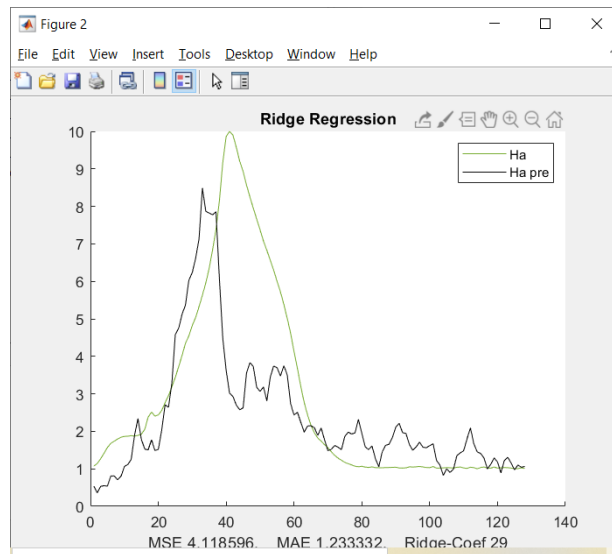


Ilustración 72. Ha Leyes de potencia Calentamiento 4

Calentamiento tipo 5

Este tipo de calentamiento se ha analizado respecto de la descarga 37596, o bien la descarga número 228 de la base de datos con la que se realiza el estudio.

Variable T_e

Mediante el análisis de la variable temperatura en el calentamiento de tipo 5, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

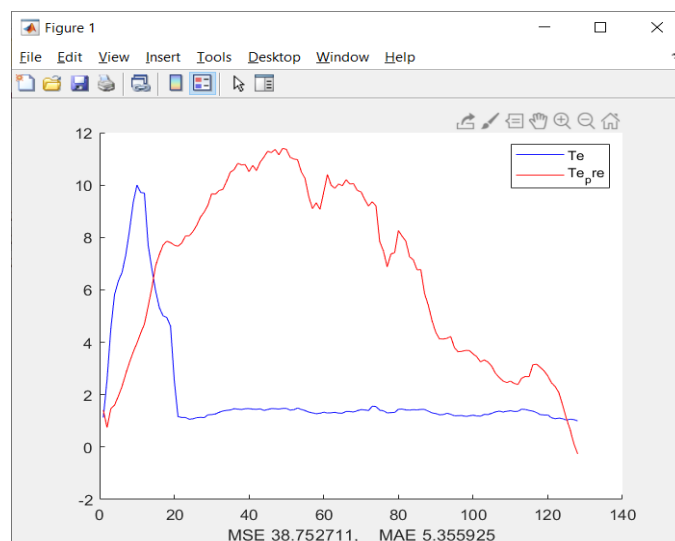


Ilustración 73. Te Regresión simbólica Calentamiento 5

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

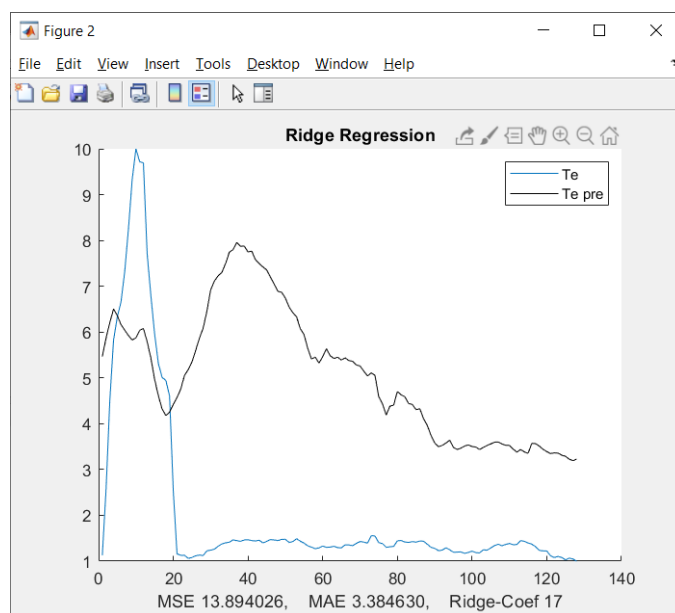


Ilustración 74. N Leyes de potencia Calentamiento 5

Variable n

Mediante el análisis de la variable densidad en el calentamiento de tipo 5, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

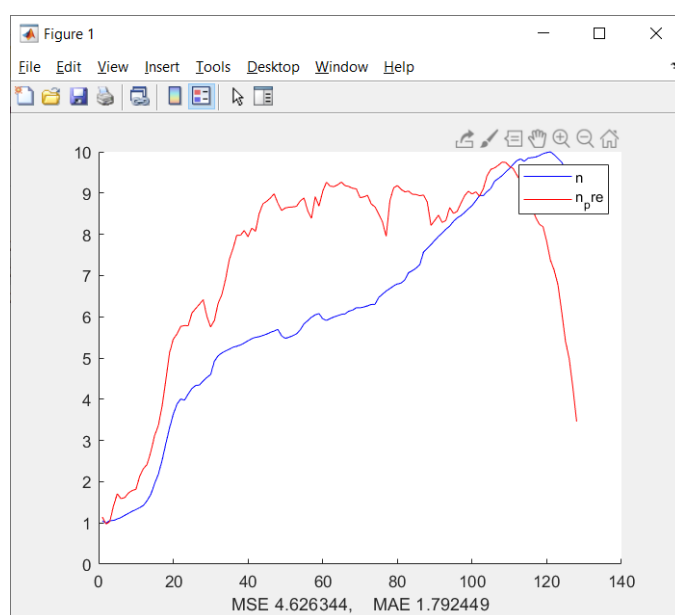


Ilustración 75. N Regresión simbólica Calentamiento 5

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Robust Regression.

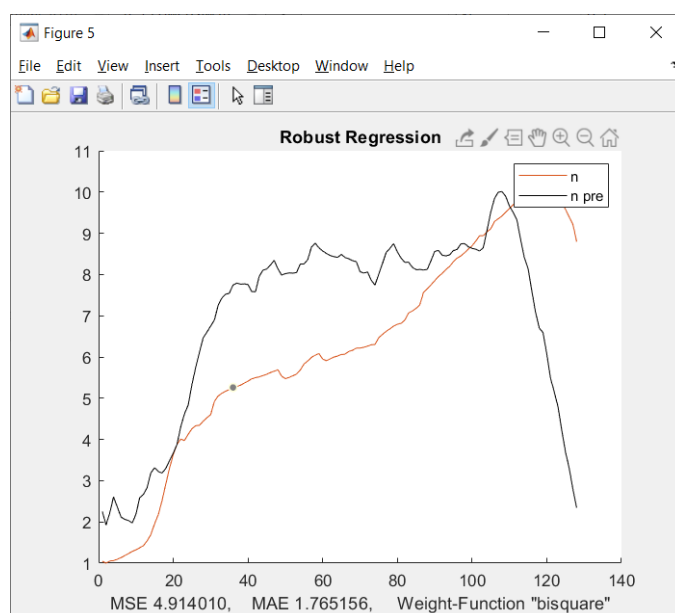


Ilustración 76. Te Leyes de potencia Calentamiento 5

Variable W_p

Mediante el análisis de la variable W_p en el calentamiento de tipo 5, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

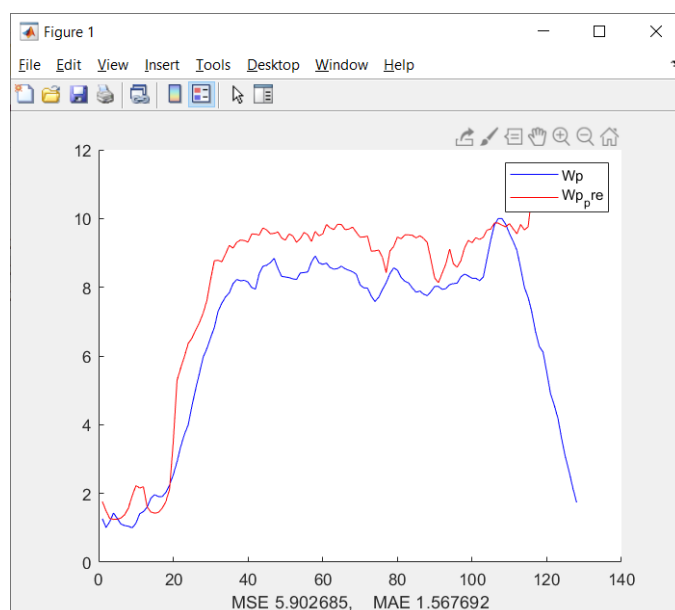


Ilustración 77. W_p Regresión simbólica Calentamiento 5

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

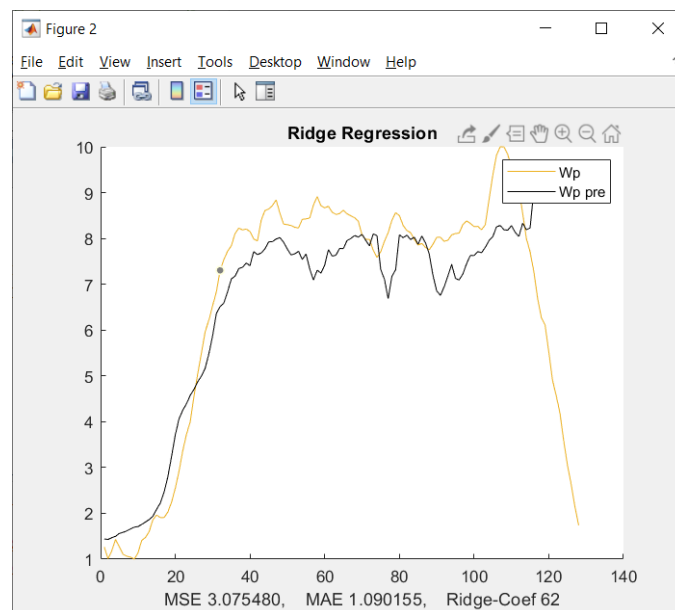


Ilustración 78. Wp Leyes de potencia Calentamiento 5

Variable I_p

Mediante el análisis de la variable I_p en el calentamiento de tipo 5, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

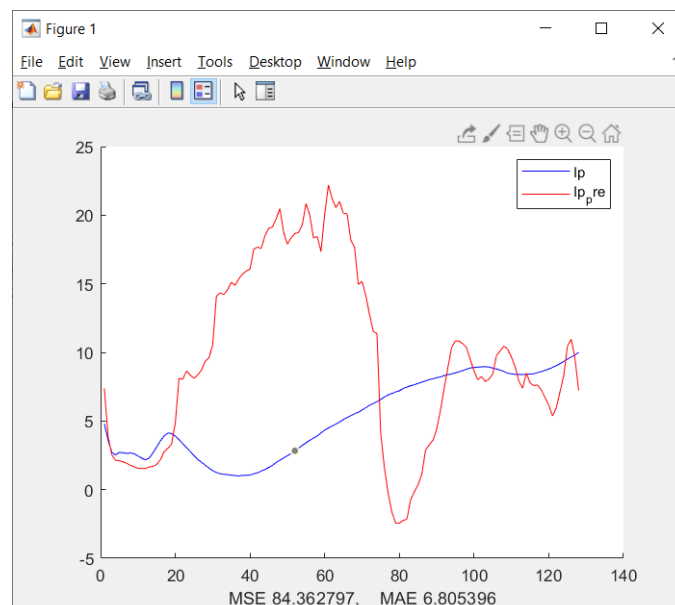


Ilustración 79. I_p Regresión simbólica Calentamiento 5

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

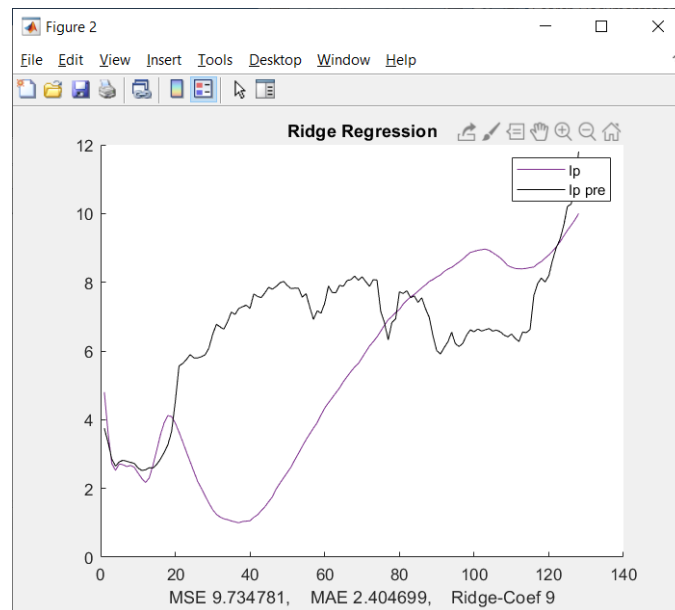


Ilustración 80. Ip Leyes de potencia Calentamiento 5

Variable Ha

Mediante el análisis de la variable H-alpha en el calentamiento de tipo 5, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

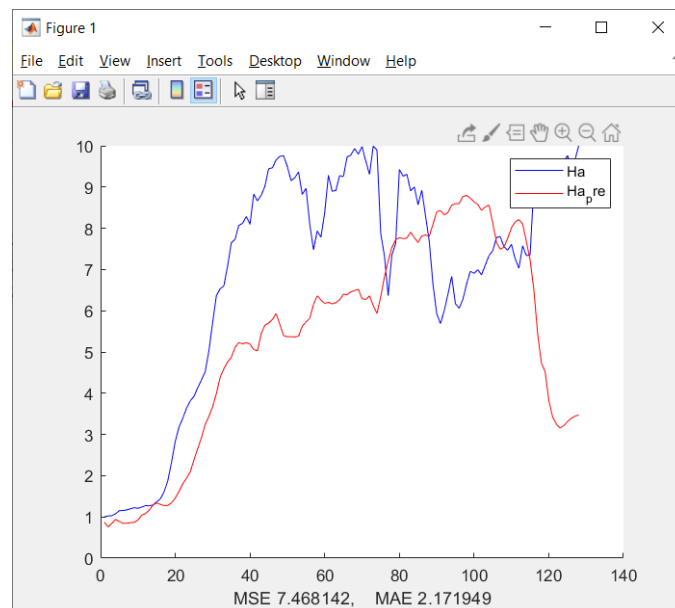


Ilustración 81. Ha Regresión simbólica Calentamiento 5

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Total Least Square.

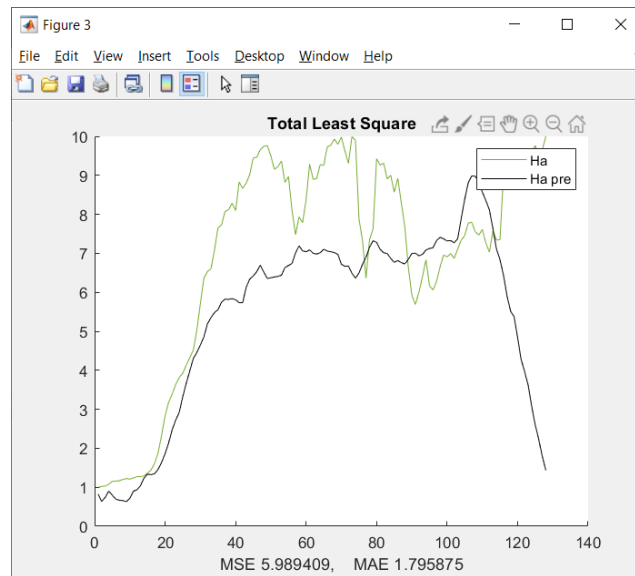


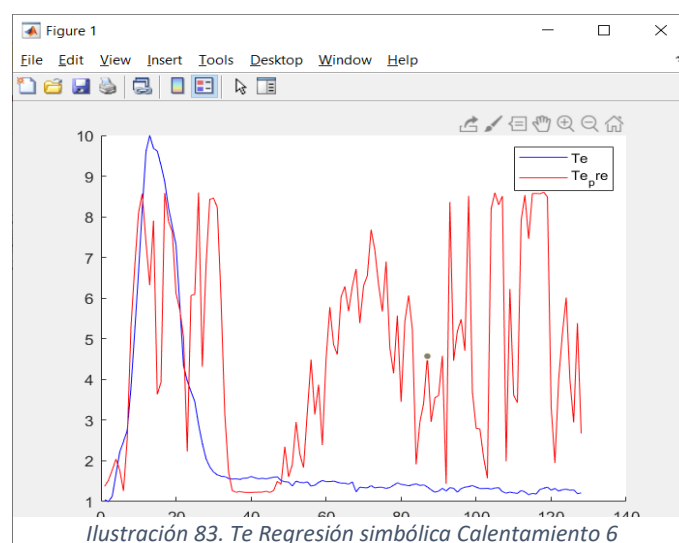
Ilustración 82. Ha Leyes de potencia Calentamiento 5

Calentamiento tipo 6

Este tipo de calentamiento se ha analizado respecto de la descarga 44902, o bien la descarga número 437 de la base de datos con la que se realiza el estudio.

Variable T_e

Mediante el análisis de la variable temperatura en el calentamiento de tipo 6, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.



A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

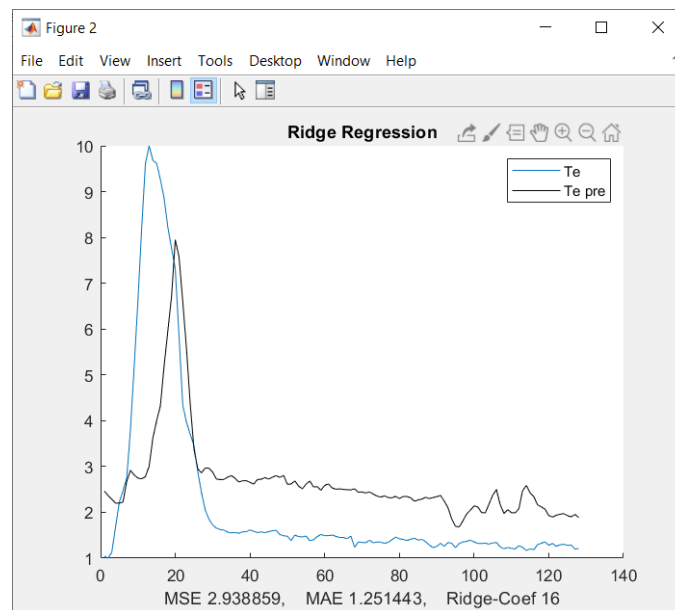


Ilustración 84. Te Leyes de potencia Calentamiento 6

Variable n

Mediante el análisis de la variable densidad en el calentamiento de tipo 6, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

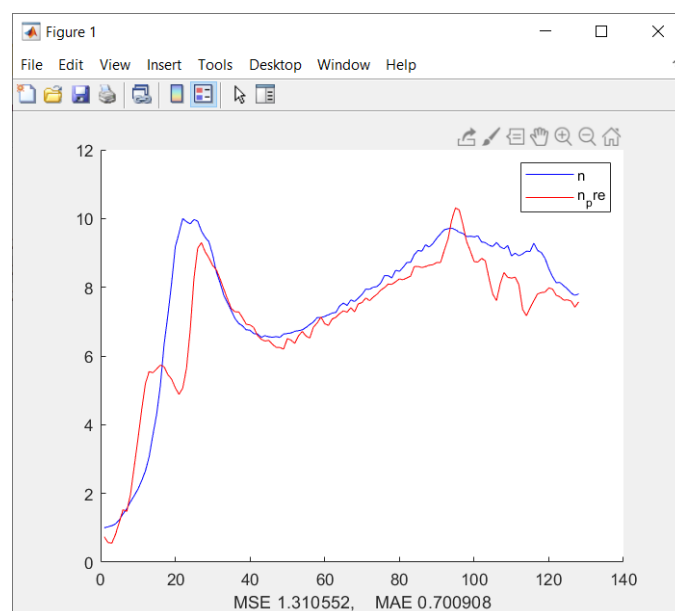


Ilustración 85. N Regresión simbólica Calentamiento 6

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

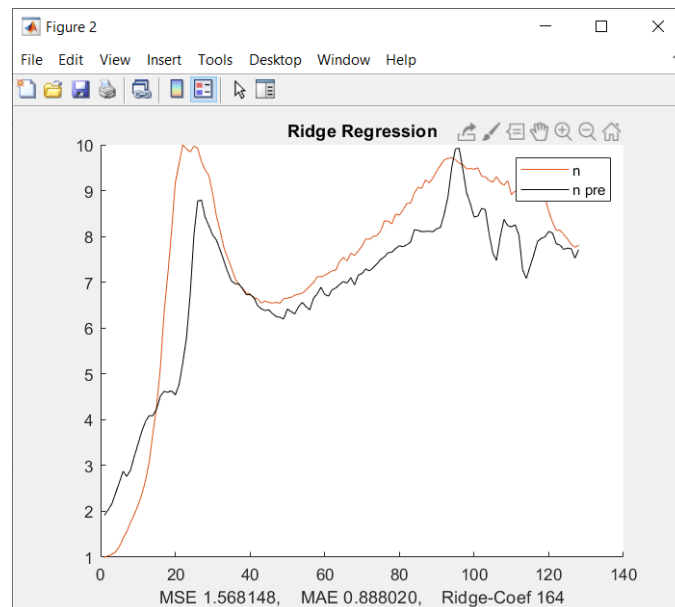


Ilustración 86. N Leyes de potencia Calentamiento 6

Variable W_p

Mediante el análisis de la variable W_p en el calentamiento de tipo 6, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

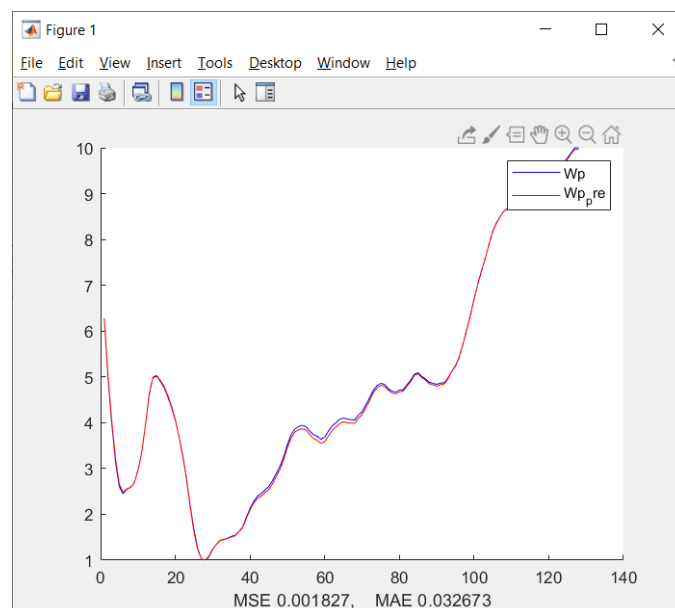


Ilustración 87. Wp Regresión simbólica Calentamiento 6

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Total Least Square.

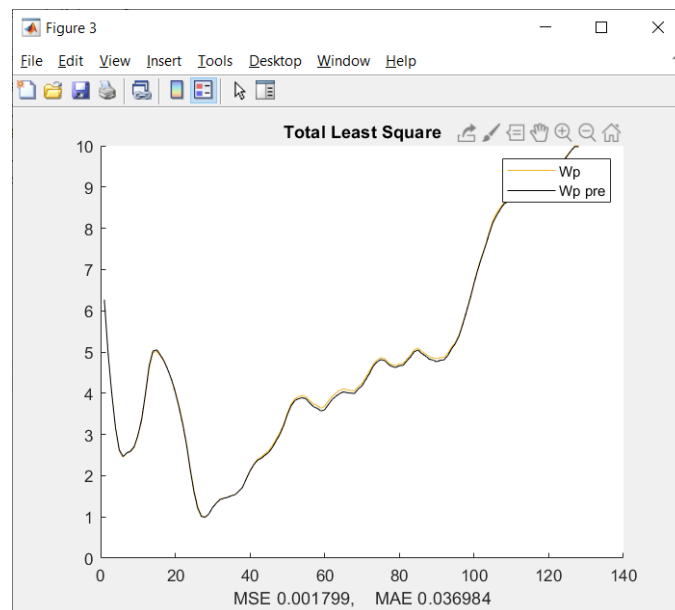


Ilustración 88. Wp Leyes de potencia Calentamiento 6

Variable I_p

Mediante el análisis de la variable I_p en el calentamiento de tipo 6, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

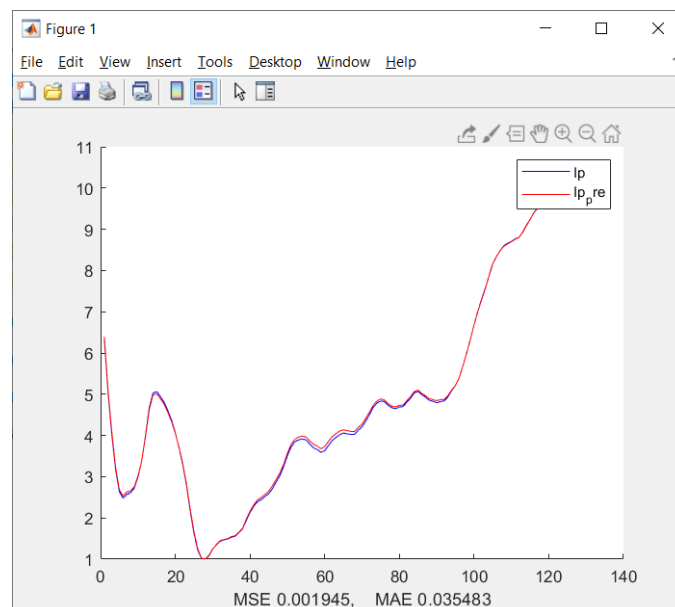


Ilustración 89. Ip Regresión simbólica Calentamiento 6

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

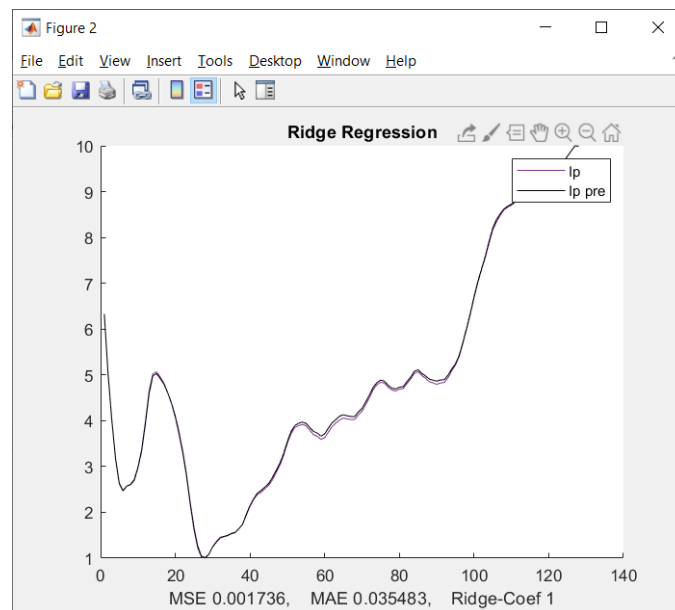


Ilustración 90. Te Leyes de potencia Calentamiento 6

Variable H_α

Mediante el análisis de la variable H-alpha en el calentamiento de tipo 6, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

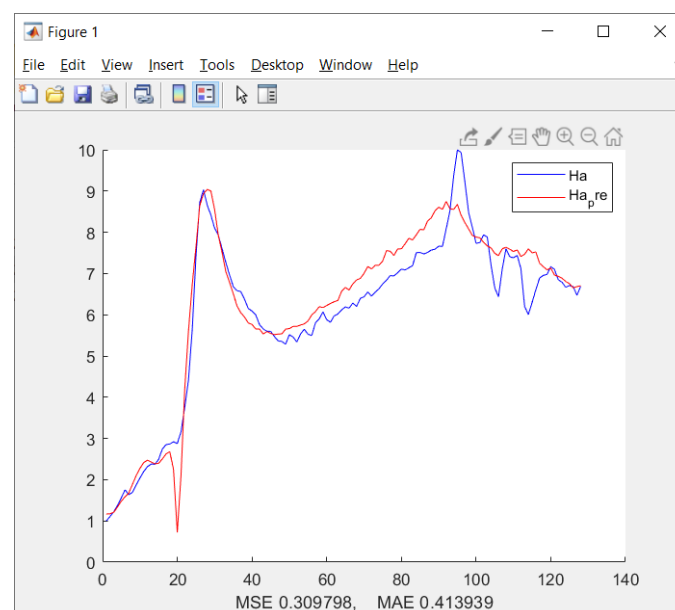


Ilustración 91. H_α Regresión simbólica Calentamiento 6

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

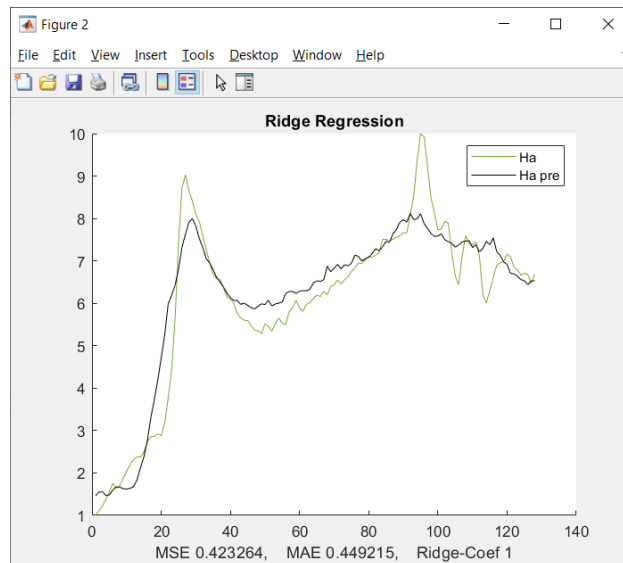


Ilustración 92. Ha Leyes de potencia Calentamiento 6

Calentamiento tipo 7

Este tipo de calentamiento se ha analizado respecto de la descarga 38899, o bien la descarga número 298 de la base de datos con la que se realiza el estudio.

Variable T_e

Mediante el análisis de la variable temperatura en el calentamiento de tipo 6, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

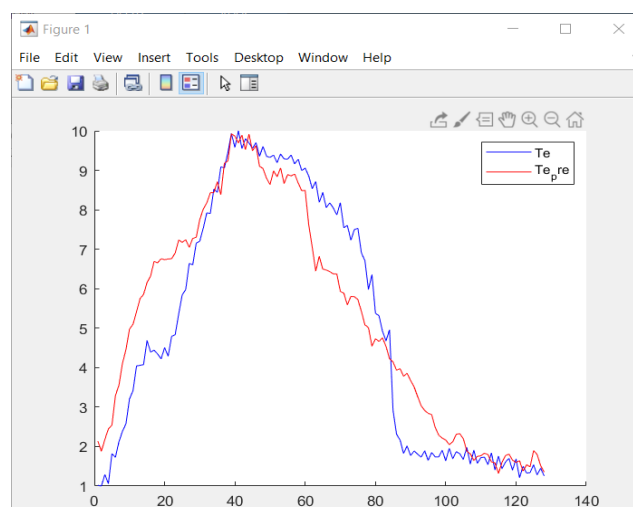


Ilustración 93. Te Regresión simbólica Calentamiento 7

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Regularized Total Least Square.

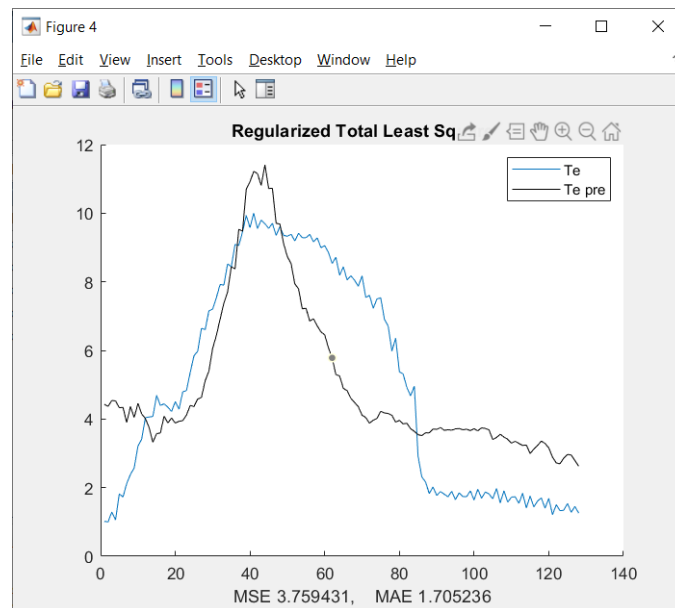


Ilustración 94. Te Leyes de potencia Calentamiento 7

Variable n

Mediante el análisis de la variable densidad en el calentamiento de tipo 6, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

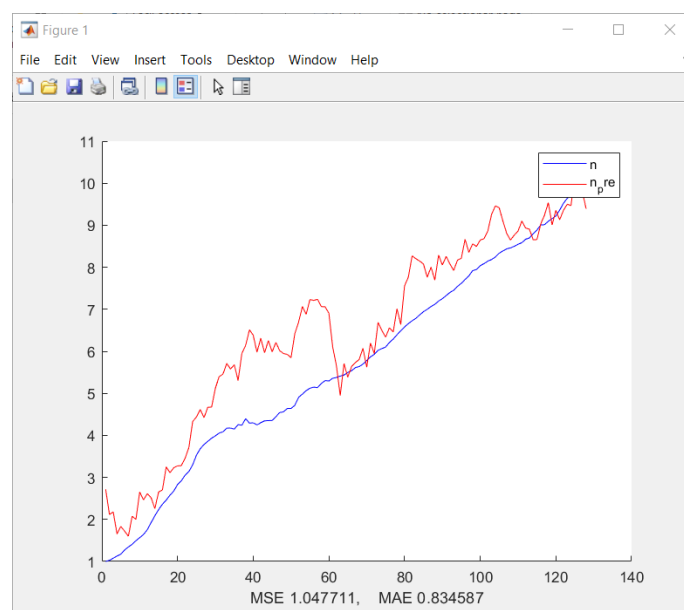


Ilustración 95. N Regresión simbólica Calentamiento 7

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

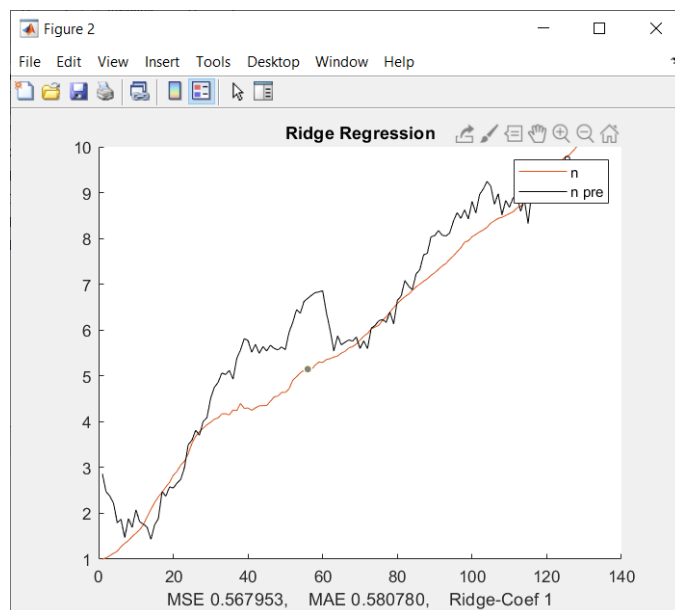


Ilustración 96. N Leyes de potencia Calentamiento 7

Variable W_p

Mediante el análisis de la variable W_p en el calentamiento de tipo 6, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

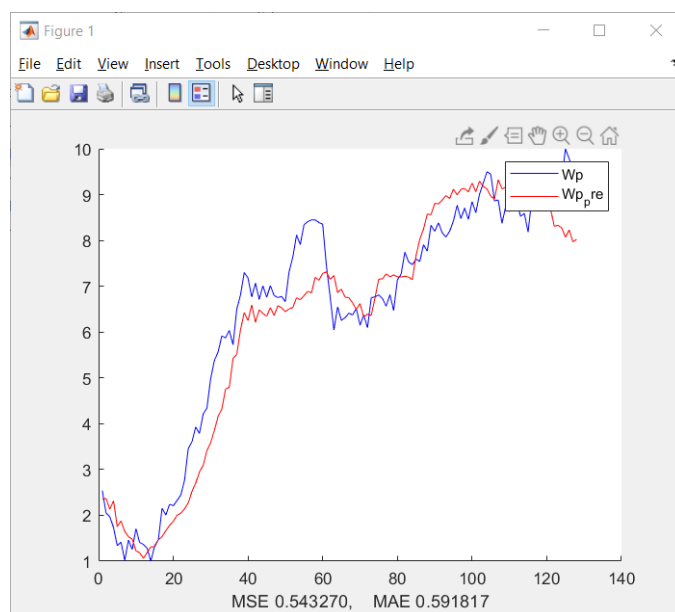


Ilustración 97. W_p Regresión simbólica Calentamiento 7

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

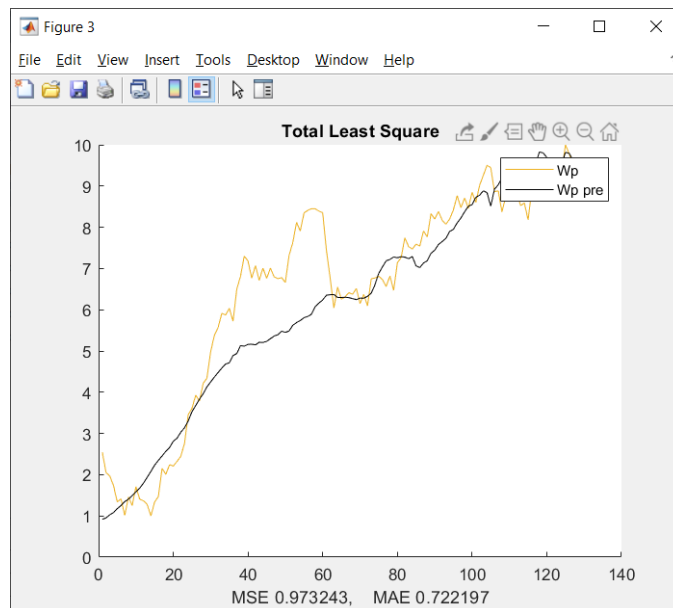


Ilustración 98. Wp Leyes de potencia Calentamiento 7

Variable I_p

Mediante el análisis de la variable I_p en el calentamiento de tipo 6, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

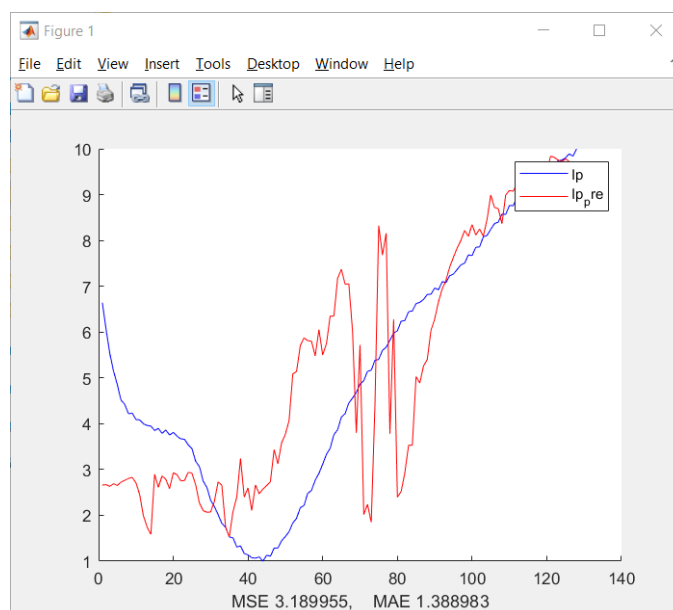


Ilustración 99. I_p Regresión simbólica Calentamiento 7

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

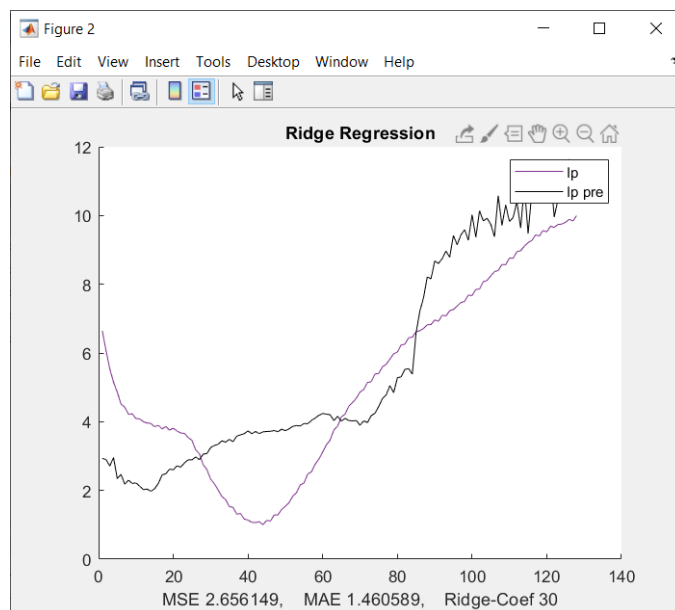


Ilustración 100. Ip Leyes de potencia Calentamiento 7

Variable Ha

Mediante el análisis de la variable H-alpha en el calentamiento de tipo 6, se pretende obtener una conclusión clara acerca de que técnica muestra una mayor precisión y exactitud con la curva real a la que deseamos asemejarnos.

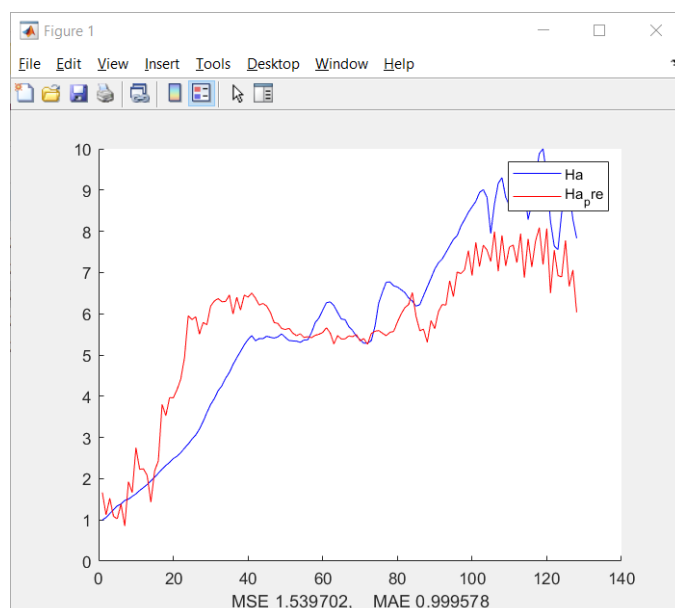


Ilustración 101. Ha Regresión simbólica Calentamiento 7

A la hora de ejecutar el script referente a las leyes de potencia en el cual se muestra las distintas técnicas empleadas, el modelo que presenta menor error es el de Ridge Regression.

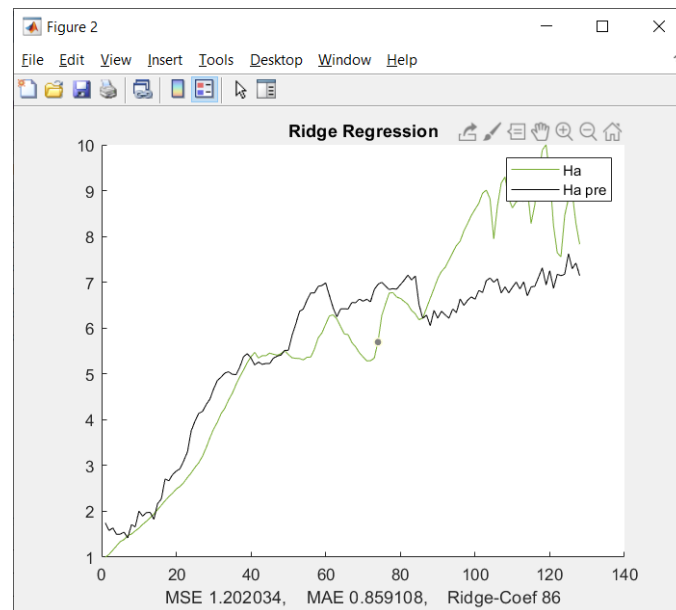


Ilustración 102. Ha Leyes de potencia Calentamiento 7

Conclusiones

Tras la realización de todas las iteraciones de código, desarrolladas mediante los softwares Eureka y Matlab, se obtienen doscientas diez (210) gráficas. De la totalidad de estas gráficas, treinta y cinco (35) proceden de la técnica de regresión simbólica, mientras que las ciento setenta y cinco (175) restantes tienen como origen la aplicación de las distintas técnicas procedentes de las leyes de potencia.

Tras observar dicha gráfica, se contempla que una gran parte del trabajo realizado se ha enfocado sobre la evaluación exhaustiva de las leyes de potencia en todas sus distintas modalidades, con el fin de comparar la regresión paramétrica y las leyes de potencia, pero también para conocer cuál de esta última es la técnica que más se suele ajustar al modelo real.

En el tipo de calentamiento 1, se puede observar cómo existe una clara superioridad del modelo de regresión simbólica frente a las distintas técnicas de potencia, donde el resultado existente es de cuatro (4) frente a uno (1), respectivamente. No obstante, la técnica que lidera el grupo pequeño es la de Ridge Regression.

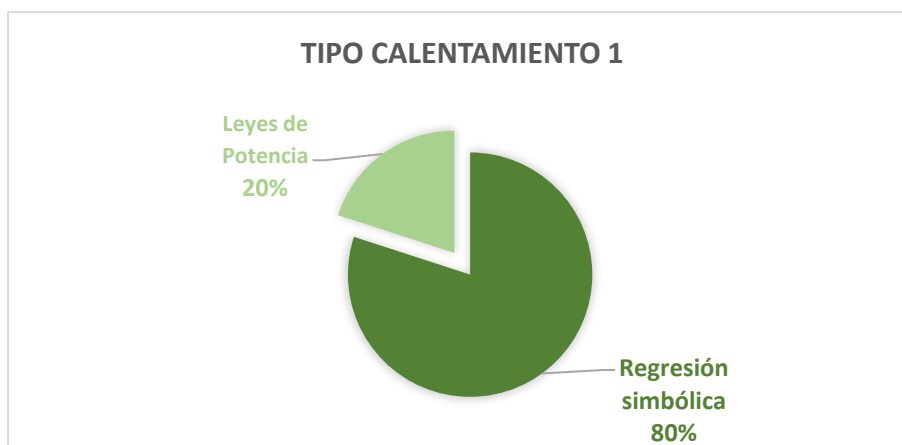


Ilustración 103.Comparativa calentamiento 1

En el tipo de calentamiento 2, se contempla una clara dominancia del modelo de leyes de potencia frente a las distintas técnicas de regresión simbólica, donde el resultado existente es de cuatro (4) frente a uno (1), respectivamente. Además, cabe destacar que la técnica que lidera el grupo grande es la de Ridge Regression y Robust Regression.

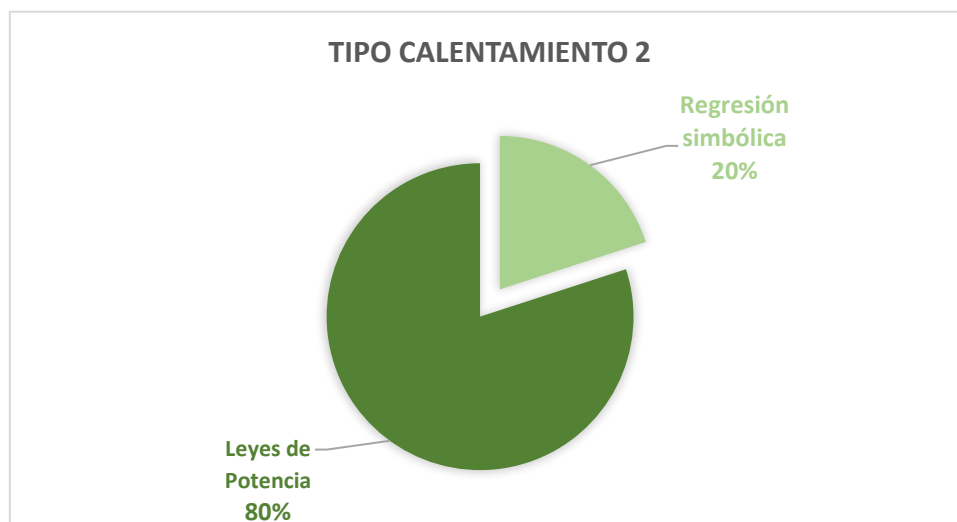


Ilustración 104. Comparativa calentamiento 2

En el tipo de calentamiento 3, se contempla una clara dominancia del modelo de leyes de potencia frente a las distintas técnicas de regresión simbólica, donde el resultado existente es de cuatro (4) frente a uno (1), respectivamente. Además, cabe destacar que la técnica que lidera el grupo grande es la de Robust Regression.

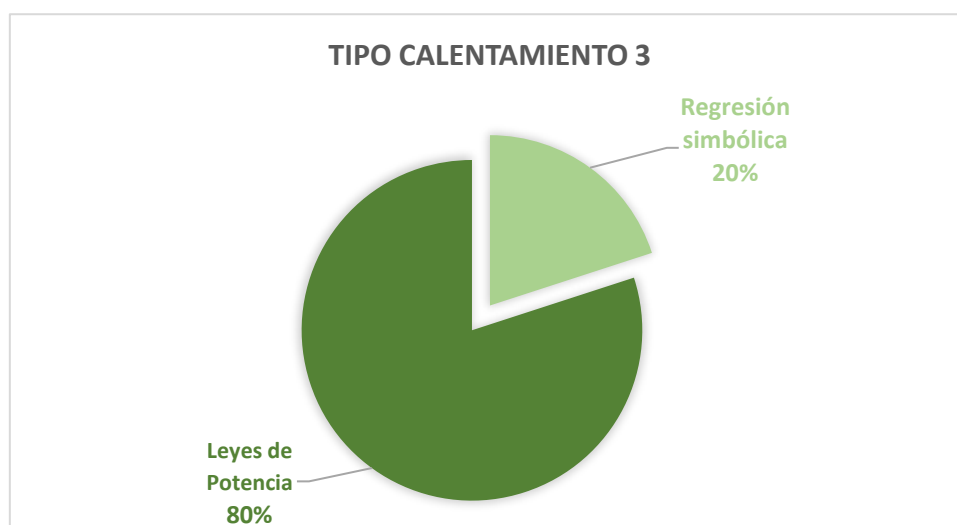


Ilustración 105. Comparativa calentamiento 3

En el tipo de calentamiento 4, se puede observar cómo existe una clara superioridad del modelo de regresión simbólica frente a las distintas técnicas de potencia, donde el resultado existente es de cuatro (4) frente a uno (1), respectivamente. No obstante, la técnica perteneciente al grupo de las técnicas de potencia empleadas que lidera el grupo es la de Total Least Square.

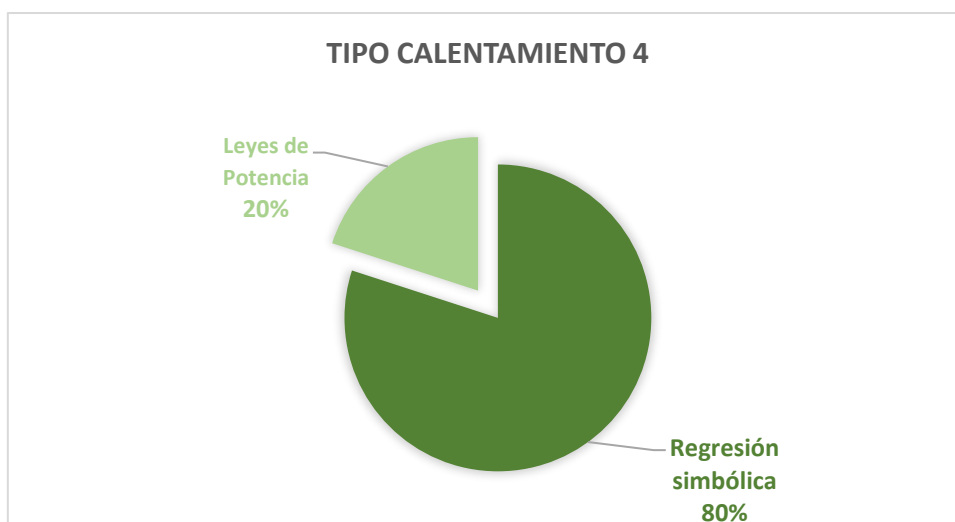


Ilustración 106. Comparativa calentamiento 4

En el tipo de calentamiento 5, se contempla una clara dominancia del modelo de leyes de potencia frente a las distintas técnicas de regresión simbólica, donde el resultado existente es de cuatro (4) frente a uno (1), respectivamente. Además, cabe destacar que la técnica que lidera el grupo grande es la de Ridge Regression.

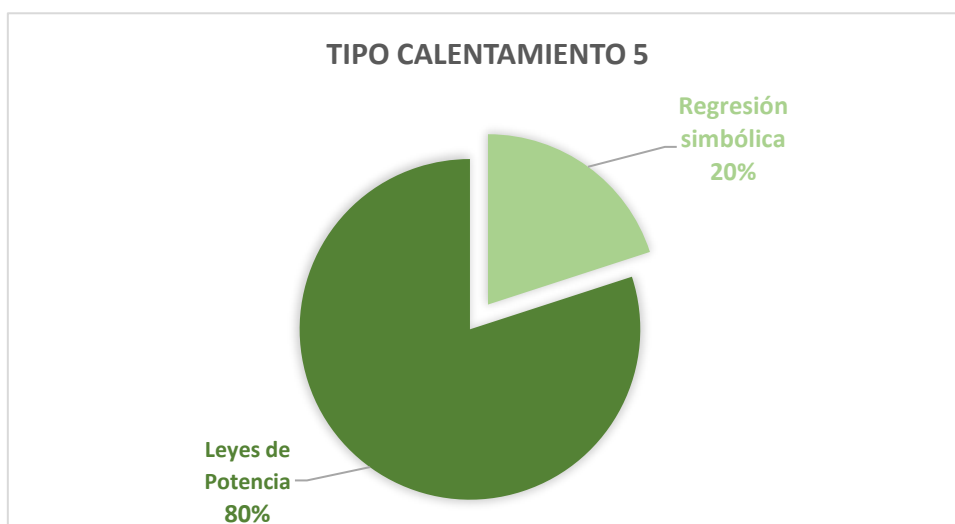


Ilustración 107. Comparativa calentamiento 5

En el tipo de calentamiento 6, se contempla una un equilibrio en cuanto a la dominancia de modelos entre las leyes de potencia y las distintas técnicas de regresión simbólica, donde el resultado existente es de dos (2) frente a tres (3), respectivamente. Además, cabe destacar que la técnica que lidera el grupo grande es la de Ridge Regression.

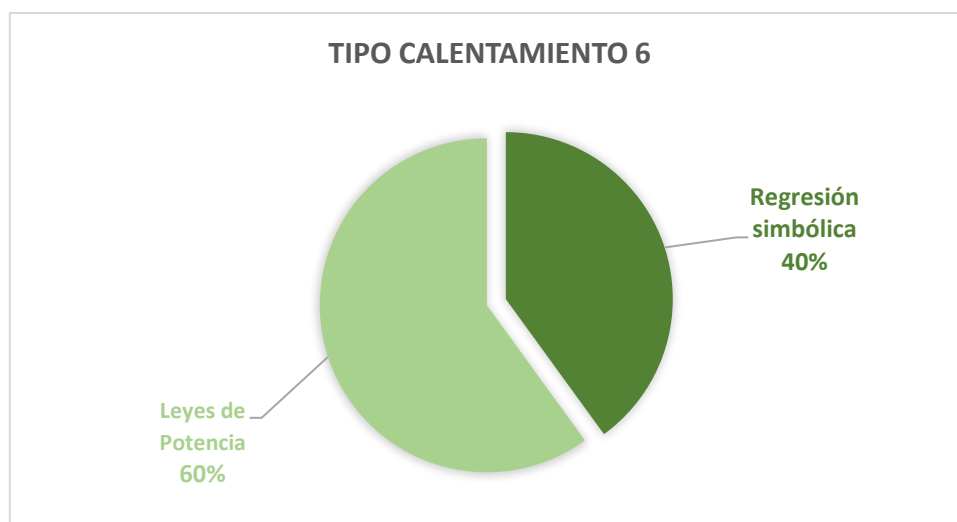


Ilustración 108. Comparativa calentamiento 6

Finalmente, en el tipo de calentamiento 7, se contempla también un equilibrio en cuanto a la dominancia de modelos entre las leyes de potencia y las distintas técnicas de regresión simbólica, donde el resultado existente es de dos (2) frente a tres (3), respectivamente. Además, la técnica que lidera el grupo más grande es la de Ridge Regression.

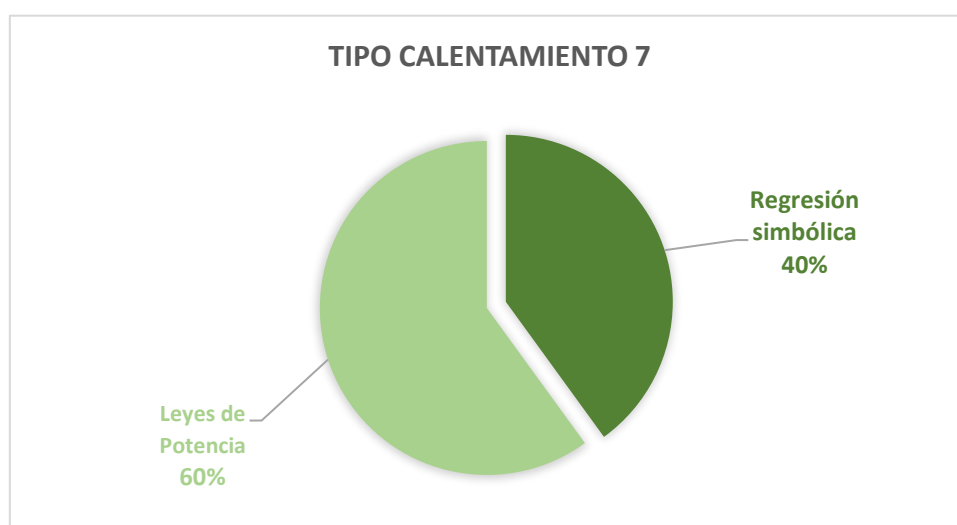


Ilustración 109. Comparativa calentamiento 7

Por lo tanto, tras realizar todas estas conclusiones, se llega a la conclusión si se suma el número de veces que ha liderado cada técnica a nivel general, se observaría que las técnicas de regresión simbólica vencen en quince (15) ocasiones, mientras que las de potencia en veinte (20). Esto supone un gran equilibrio entre ambos procedimientos, dado que no existe una superioridad absoluta de una frente a la otra. De este modo, los porcentajes de dominancia resultarían en

43% y 57% respectivamente, aunque a pesar de ello, el método de potencia muestra una mucha mayor exactitud en los casos que lidera, reduciendo el error en ocasiones a un tercio.

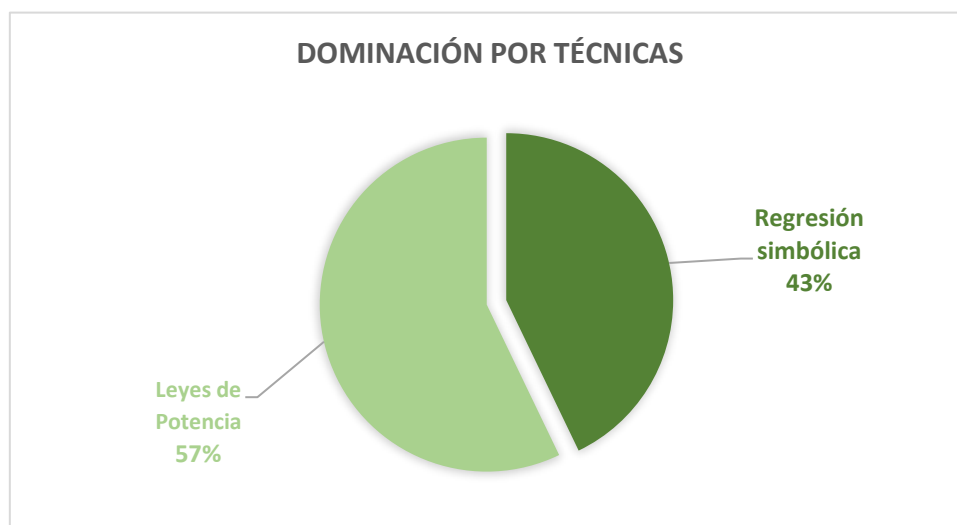


Ilustración 110. Dominación por gráficas

A su vez, se ha de realizar una mención acerca de que el modelo perteneciente a la metodología de potencia que ha dominado internamente entre todos los posibles procedimientos de dicha sección de ensayos, frente al resto ha sido la técnica de Ridge Regression, donde esta logra imponerse en un total de doce (12) ocasiones frente a las 20 del total en las se impone. Y esto implica que se dan casos de colinealidad entre las distintas variables independientes que definen los modelos matemáticos, y como consecuencia pueden existir errores mayores. Sin embargo, dicho proceso se encarga de añadir un grado más de seguridad en las estimaciones relativas que se hacen, por lo que entonces esto implica que los cálculos que se suelen llevar a cabo, presentan patrones internos muy relacionados entre sí por lo que permite demostrar que el estudio realizado es correcto, y tiene un gran trasfondo a nivel teórico-práctico, ya que se ha conseguido obtener un modelo matemático capaz de hacer frente a una gran parte de las situaciones.

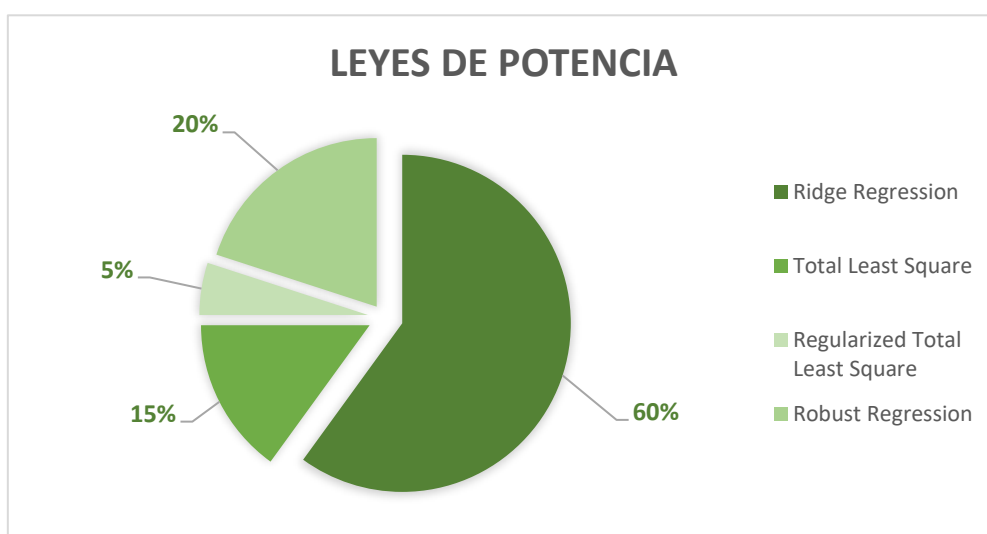


Ilustración 111. Orden de logros de Leyes de potencia

No obstante, se ha de hacer mención en que otra conclusión a la que se ha podido llegar es que a medida que la base de datos referente a dicho tipo de calentamiento es mucho mayor, véase el caso del calentamiento tipo 1, que presenta más de doscientas (200) descargas, o el calentamiento tipo 4 con más de cien (100) descargas, se muestra un error mucho más pequeño en todas sus gráficas y además el modelo matemático que se aproxima más a la realidad es el procedente de las técnicas de regresión simbólica.

Sin embargo, por el otro lado en los tipos de calentamiento en los cuales el número de descargas es mucho menor, y por tanto la base de datos que presenta más puntos en común con dicha descarga a analizar se ve más reducida, lo que ocurre es que los modelos que se encargan de diseñar mejor el modelo que se aproxima a la realidad es el referente a leyes de potencia, logrando reducir el error en gran medida en comparación con la programación genética.

De este modo, a modo aclaratorio los resultados obtenidos y comparados en función de cuál es el método que más se ha aproximado a la curva real, quedaría reflejado del siguiente modo.

	Temperatura	Densidad	Energía Diamagnética	Corriente del plasma	H-alfa
Calentamiento 1	Regresión simbólica	Regresión simbólica	Regresión simbólica	Ridge Regression	Regresión simbólica
Calentamiento 2	Regresión simbólica	Robust Regression	Ridge Regression	Ridge Regression	Robust Regression
Calentamiento 3	Robust Regression	Robust Regression	Regresión simbólica	Ridge Regression	Regularized Total Least Square
Calentamiento 4	Total Least Square	Regresión simbólica	Regresión simbólica	Regresión simbólica	Regresión simbólica
Calentamiento 5	Ridge Regression	Regresión simbólica	Ridge Regression	Ridge Regression	Total Least Square
Calentamiento 6	Ridge Regression	Regresión simbólica	Total Least Square	Ridge Regression	Regresión simbólica
Calentamiento 7	Regresión simbólica	Ridge Regression	Regresión simbólica	Ridge Regression	Ridge Regression

Tabla 1. Resultados del análisis comparativo

Bibliografía

[A.Martínez et al,2013] A.Martínez, D.Velásquez-Henao. **Una modificación de la metodología de regresión simbólica para la predicción de series de tiempo.** Towards Data Science.

<http://www.scielo.org.co/pdf/inun/v17n2/v17n2a04.pdf>

[Bannai,2018] Ryota Bannai. **Total Least Squares in comparison with OLS and ODR.** Tesis de maestría: Problemas abiertos en la aplicación de la regresión simbólica en el pronóstico de series de tiempo por C. A. Martínez.

<https://towardsdatascience.com/total-least-squares-in-comparison-with-ols-and-odr-f050ffc1a86a>

[Barker,1999] Barker, J. (1999). **Ley de potencia.** Nist Center for Neutron Research.

<https://www.ncnr.nist.gov/resources/sansmodels/PowerLaw.html>

[Brownlee,2020] Jason Brownlee. **Robust Regression for Machine Learning in Python.** Machine Learning Mastery.

<https://machinelearningmastery.com/robust-regression-for-machine-learning-in-python/>

[Cusiritati,2020] Cusiritati. **Las ventajas y desventajas de un modelo de regresión múltiple.**

<https://www.cusiritati.com/EzQavp2zg/>

[Dudley,2012] Dudley L. Poston Jr. **Ordinary Least Squares Regression.** International Encyclopedia of the Social Sciences. Encyclopedia.com.

<https://www.encyclopedia.com/social-sciences/applied-and-social-sciences-magazines/ordinary-least-squares-regression>

[Florido,2018] Florido Florido, F. **Librería para Algoritmos Genéticos Basados en Nichos.** Trabajo Fin de carrera Universidad Carlos III 2015.

https://e-archivo.uc3m.es/bitstream/handle/10016/25867/PFC_FranciscoJavier_Florido_Florido.pdf

[Gauss,1809] Gauss, Carl Friedrich. **Teoría del movimiento de los cuerpos celestes que se mueven alrededor del sol en secciones cónicas.** Editado por Sociedad de Ciencias de Gotinga (1870-1871).

[Glen,2016] Glen, Stephanie. **Power Law and Power Law Distribution.** Statistics How To.

<https://www.statisticshowto.com/power-law/>

[Golub et al,1970] Golub and Reinsch. **Singular Value Decomposition and Least Squares Solutions.** Springer Link. Numerische Mathematik 14, 403--420

<http://people.duke.edu/~hpgavin/SystemID/References/Golub%2BReinsch-NM-1970.pdf>

[Golub et al,1980] G.H. Golub, C.F. Van Loan. **An analysis of the total least squares problem.** SIAM J. Numerische Mathematik, 17, 883-893.

[Golub et al,1980b] G.H. Golub, P.C. Hansen, D. O'Leary. **Tikhonov Regularization and Total Least Squares.** SIAM J. Matriz Analitical Application. Vol. 21, No. 1, 185-194.

[Gratton et al,2013] Serge Gratton , David Titley-Peloquin y Jean Tshimanga Ilunga. **Sensibilidad y acondicionamiento de la solución de mínimos cuadrados totales**

truncados. Revista SIAM sobre análisis de matrices y aplicaciones. Science Comput., 18(4), 1223–1241.

[Guo et al,2008] Hongbin Guo, Rosemary A. Renaut. **A Regularized Total Least Squares Algorithm.** Springer. ISBN: 978-94-017-3552-0
https://doi.org/10.1007/978-94-017-3552-0_6

[Hamilton,1992] Hamilton, Lawrence C. 1992. **Regresión con gráficos: un segundo curso en estadística aplicada.** Pacific Grove, CA: Brooks / Cole.

[Hoerl et al,1970] Arthur E. Hoerl, Robert W. Kennard. **Applications to Nonorthogonal Problems.** Technometrics, Vol. 12, No. 1., pp. 69-82.
<https://www.jstor.org/stable/1267352>

[Holland,1977] Holland, P. W., and R. E. Welsch. **Robust Regression Using Iteratively Reweighted Least-Squares.** Communications in Statistics: Theory and Methods, A6, 1977, pp. 813–827.

[ITER,2015] ITER y Foro Nuclear. **El proyecto de fusión nuclear ITER.** ForoNuclear.org.
<https://www.foronuclear.org/actualidad/a-fondo/el-proyecto-de-fusion-nuclear-iter/>

[Lawson et al,1987] Lawson y Hanson. **Total Least Squares.** Society for Industrial and Applied Mathematics. Appendix A, pág. 2.
<https://link.springer.com/content/pdf/bbm%3A978-3-319-89620-5%2F1.pdf>

[Lewis-Beck et al,1980] Lewis-Beck, Michael S. 1980. **Regresión aplicada: Introducción.** SAGE Publications, N.º 2 edición (2015).

[Moreno,2008] Moreno Parra, R. **Programación Genética: La Regresión Simbólica.** Artículo vinculado a página web Darwin.50webs.
<http://armando2k.com/coorelacion.pdf>

[NCSS,2012] NCSS Statistical Software. **Ridge Regression.** Chapter 335.
https://ncss-Wpengine.netdna-ssl.com/Wp-content/themes/ncss/pdf/Procedures/NCSS/Ridge_Regression.pdf

[Olivera,2017] Nahuel Olivera Rodriguez. **Wavelets de Haar y Daubechies y sus aplicaciones.** Trabajo Fin de Grado en Universidad Jaume I. 2018.
http://repositori.uji.es/xmlui/bitstream/handle/10234/177642/TFG_Olivera%20Rodriguez%2C%20Nahuel.pdf?sequence=1&isAllowed=y

[Pereira,2010] Augusto Pereira González. **Análisis Predictivo de datos mediante técnicas de regresión estadística.** Trabajo Fin de Máster en Universidad Complutense de Madrid. 2010
https://eprints.ucm.es/id/eprint/11389/1/Analisis_Predictivo_de_Datos.pdf

[Pereira,2015] Augusto Pereira González. **Selección de características para el reconocimiento de patrones con datos de alta dimensionalidad en fusión nuclear.** Tesis doctoral en Universidad Nacional Española a Distancia. 2015.
http://e-spacio.uned.es/fez/eserv/tesisuned:IngInf-Apereira/PEREIRA_GONZALEZ_Augusto_Tesis.pdf

[Pértega et al,2000] Pértega Díaz, S., Pita Fernández, S. **Técnicas de regresión: Regresión Lineal Múltiple**. Fistera. Cad Aten. Primaria 2000; 7: 173-176.

<https://www.fistera.com/formacion/metodologia-investigacion/tecnicas-regresion-regresion-lineal-multiple/>

[Poston,2002] Poston, Dudley L., Jr. 2002. **El modelo estadístico de la fertilidad de las mujeres chinas**. Revista de métodos estadísticos aplicados modernos 1 (2): 387 - 396.

<http://www.ub.edu/geocrit/sn/sn-348.htm>

[Poston et al,2003] Poston, Dudley L., Jr. y Sherry L. McKibben. 2003. **Modelos de regresión de conteo inflados con cero para estimar la fertilidad de las mujeres estadounidenses**. Revista de métodos estadísticos aplicados modernos 2 (2): 371 – 379.

[Powerdata,2020] Powerdata. **Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad**. Centro de recursos de Powerdata.

<https://www.powerdata.es/big-data>

[Schmidt,2009] Michael Schmidt. **Distilling Free-Form Natural Laws From Experimental Data**. Science 324, 81. Vol. 324, Issue 5923, 81-85. DOI: 10.1126/science.1165893

<https://science.sciencemag.org/content/324/5923/81.full>

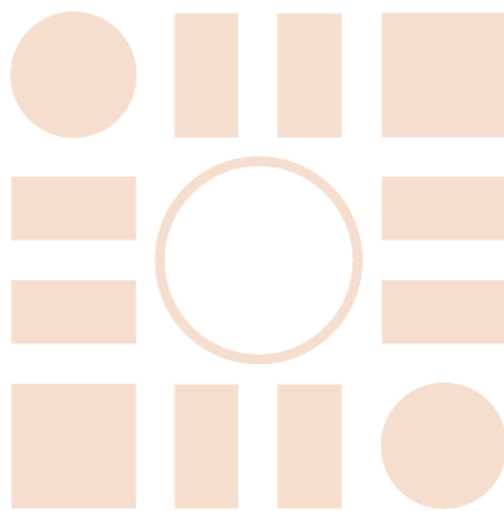
[Smidt, et al., 2013] Eureka (version 1.24.0 (build 9367) Software disponible y bajado como último acceso en Noviembre de 2015 y utilizada solamente en el periodo de prueba disponible y vigente desde su instalación.

www.creativemachineslab.com/eureqa.html

[UCLA,2019] Institute for Digital Research & Education Statistical Consulting UCLA. **ROBUST REGRESSION | R DATA ANALYSIS EXAMPLES**. Artículo de UCLA (2015).

<https://stats.idre.ucla.edu/r/dae/robust-regression/>

Universidad de Alcalá
Escuela Politécnica Superior



ESCUELA POLITECNICA
SUPERIOR



Universidad
de Alcalá